

PENGEMBANGAN TES MENGGUNAKAN MODEL RASCH MATERI GAYA UNTUK SMA

ARTIKEL PENELITIAN



**OLEH:
MUHAMAD SAFIHIN
NIM. F1051141017**

**PROGRAM STUDI PENDIDIKAN FISIKA
JURUSAN PENDIDIKAN MATEMATIKA DAN IPA
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
PONTIANAK
2019**

LEMBAR PERSETUJUAN

PENGEMBANGAN TES MENGGUNAKAN MODEL RASCH MATERI GAYA UNTUK SMA

ARTIKEL PENELITIAN

MUHAMAD SAFIHIN
NIM .F1051141017

Disetujui,

Pembimbing I

Dr. Haratua Tiur Maria S, M.Pd

NIP. 196702221991012001

Pembimbing II

Hamdani, M.Pd

NIP. 19198506052008121001

Mengetahui,

Ketua Jurusan P. MIPA

Dr. Ahmad Yani T, M.Pd

NIP. 1966040119910210001



PENGEMBANGAN TES MENGGUNAKAN MODEL RASCH MATERI GAYA UNTUK SMA

Muhamad Safihin, Haratua Tiur Maria S., Hamdani

Program Studi Guru Fisika FKIP Untan Pontianak

Email: safihinhaikal@gmail.com

Abstract

This research aimed to produce an objective test in the form of multiple choice test in the Force subject to measure the learning outcomes of the tenth grade students of Senior High School in Pontianak City. This study uses the Research and Development (R & D) method with the Four-D model. The development of the test used the Mardapi steps consisting of 9 steps. The initial design of the test consisted of 40 items of multiple choice test. The trial test involved 34 students in a small scale trial, and 202 students in a large-scale. The characteristics of the test items were analyzed using the Winstep program with the Rasch Model. Based on the data analysis, it was concluded that the test developed could be used to measure student learning outcomes, namely: 1) items distributed evenly on the person-item map, 2) having difficulty levels varied from easy to difficult as many as 25 items, 3) unidimensionality is more than minimum requirements of 20%, that is equal to 26.2%, 4) sufficient reliability, which is equal to 0.68, 5) questions are made in simple language with a readability level of 7.08. The questions are suitable to be used are 25 items.

Keywords: Force Subject, Rasch Model, Test Development.

PENDAHULUAN

Penilaian merupakan bagian yang penting dalam proses pembelajaran. Penilaian adalah penerapan berbagai cara dan penggunaan beragam alat penilaian untuk memperoleh informasi tentang sejauh mana hasil belajar peserta didik atau ketercapaian kompetensi peserta didik (Depdiknas, 2004). Dengan melakukan penilaian, Guru sebagai pengelola kegiatan pembelajaran dapat mengetahui kemampuan yang dimiliki peserta didik, ketepatan metode mengajar yang digunakan, dan keberhasilan peserta didik dalam meraih kompetensi yang telah ditetapkan. Hal ini dipertegas dalam Permendikbud Nomor 23 Tahun 2016 tentang Standar Penilaian Guru yang menyatakan bahwa “penilaian hasil belajar oleh Guru bertujuan untuk memantau dan mengevaluasi proses, kemajuan belajar, dan perbaikan hasil belajar peserta didik secara berkesinambungan”.

Penilaian (asesmen) merupakan bagian dari evaluasi pencapaian peserta didik yang

dilakukan Guru untuk melihat hasil belajar peserta didik. Salah satu bentuk instrumen dalam penilaian kognitif yang telah dikenal dan paling sering digunakan di sekolah adalah tes. Tes ini didesain untuk mengukur hasil pembelajaran peserta didik.

Tes prestasi belajar atau (*achievement test*) merupakan salah satu bentuk tes penguasaan (*mastering test*) yang diujikan untuk mengukur penguasaan peserta didik terhadap materi yang disampaikan oleh Guru dan dipelajari oleh peserta didik (Purwanto, 2008). Dalam tes ini peserta didik didorong untuk menampilkan kemampuannya secara maksimal dan dari penampilannya dapat diketahui penguasaan peserta didik terhadap materi. Tes prestasi belajar yang digunakan dapat berupa tes yang telah distandarkan (*standardized test*) maupun tes buatan Guru sendiri (*teachermade test* atau *informal test*) (Arifin, 2009). Tes buatan Guru adalah tes hasil belajar yang disusun sendiri oleh Guru yang akan mempergunakan tes tersebut. Tes

ini biasanya digunakan untuk ulangan harian, formatif, dan ulangan umum (sumatif).

Item tes pilihan ganda merupakan jenis tes objektif yang paling banyak digunakan oleh Guru dalam melaksanakan tes hasil belajar. Hasil prasurvey di beberapa SMA Negeri dan Swasta di Kota Pontianak menunjukkan bahwa bentuk tes pilihan ganda merupakan bentuk tes yang paling sering digunakan Guru untuk mengukur kemampuan kognitif peserta didik, termasuk pada mata pelajaran Fisika

Instrumen tes pilihan ganda yang disusun haruslah memenuhi kriteria sebagai alat ukur yang baik agar dapat memberikan gambaran tentang kemampuan maupun kompetensi yang dimiliki peserta didik. Namun, selama ini kenyataan dilapangan menunjukkan instrumen tes belum mampu memberikan pengukuran yang akurat terhadap hasil belajar peserta didik. Berdasarkan hasil wawancara dengan tenaga Guru SMA di Kota Pontianak, instrumen tes yang dibuat kadang kala tidak melalui proses penyusunan instrumen tes yang benar. Hal ini dikarenakan waktu yang disediakan sekolah untuk membuat dan menyusun instrumen tes tersebut sangat singkat. Dengan waktu yang singkat dan harus memenuhi target yang ditentukan oleh sekolah menyebabkan instrumen tes buatan Guru ini tidak terstandar karena tidak didahului ujicoba, butir belum terkalibrasi, serta aspek validitas dan reliabilitas yang belum diketahui. Oleh karena itu, perlu dikembangkan instrumen tes pilihan ganda yang layak digunakan untuk mengukur hasil belajar peserta didik agar dapat mencerminkan kemampuan peserta didik yang sebenarnya.

Materi tes pilihan ganda yang dikembangkan pada penelitian ini adalah materi Gaya. Pemilihan materi ini karena Gaya merupakan konsep dasar pada pokok bahasan Dinamika Gerak. Selain itu, selama ini untuk ulangan harian pada materi Gaya, Guru disekolah lebih banyak menggunakan bentuk tes uraian singkat, dikarenakan lebih mudah dalam penulisan soal dan lebih mudah menilai kemampuan peserta didik. Namun, bentuk tes uraian singkat ini memiliki

kelemahan diantaranya memerlukan waktu yang lama dalam pengoreksia. Sehingga dirasa penting untuk mengembangkan instrumen tes pada materi Gaya ini dalam bentuk tes pilihan ganda.

Untuk menguji apakah butir – butir item yang membangun tes hasil belajar sudah dapat menjalankan fungsinya sebagai alat pengukur yang baik maka perlu dilaksanakan analisis butir soal (Sudijono, 2012). Kegiatan menganalisis butir soal merupakan suatu kegiatan yang harus dilakukan Guru untuk meningkatkan mutu butir soal yang ditulis. Dari hasil analisis tersebut, pada akhirnya akan mencerminkan karakteristik yang dimiliki oleh perangkat tes itu sendiri. Oleh karena itu, perlu dilakukan analisis butir soal untuk mengetahui seberapa layak instrumen tes yang dikembangkan tersebut dapat digunakan untuk mengukur hasil belajar peserta didik dengan akurat.

Dalam pengukuran, terdapat dua pendekatan yang sering digunakan dalam menganalisis butir soal yaitu teori tes klasik (*Classical Test Theory*) dan teori respon butir (*Item Response Theory*). Teori tes klasik memiliki sifat *group dependent* dan *item dependent*. Hal ini berarti indeks daya pembeda, tingkat kesulitan, dan koefisien reliabilitas tes tergantung kepada yang mengerjakan tes tersebut, selain dipengaruhi oleh soal atau butir yang ada (Samsul Hadi, 2013). Sedangkan dengan teori respon butir ukuran taraf kesukaran butir serta ciri butir lainnya akan tetap (invarian) terhadap kelompok peserta tes, tidak masalah kelompok peserta mana saja yang mengerjakannya selama mereka mempunyai kemampuan yang memadai untuk mengerjakannya; dan dengan teori ini pula, ukuran kemampuan peserta akan tetap (invarian) terhadap kelompok butir tes, tidak masalah kelompok butir mana saja yang mereka kerjakan selama kelompok butir itu mampu secara memadai dikerjakan oleh peserta tes (Dali S.Naga, 1992).

Bila menggunakan teori tes klasik, umumnya peserta didik menjawab butir soal suatu tes yang berbentuk pilihan ganda dengan benar diberi skor 1 dan 0 jika salah,

sehingga kemampuan peserta didik dinyatakan dengan skor total yang diperolehnya. Teori ini berasumsi semakin banyak butir soal yang dijawab benar oleh peserta didik, semakin tinggi pula kemampuan peserta didik tersebut. Prosedur tersebut kurang memperhatikan interaksi antara setiap peserta didik dengan butir soal, sehingga sulit untuk mengukur kemampuan peserta didik yang sesungguhnya.

Selain itu, menurut Hambleton & Swaminathan (1985), teori tes klasik memiliki beberapa kelemahan yaitu: (1) tingkat kesukaran dan daya beda butir soal tergantung pada kelompok peserta yang mengerjakannya, (2) penggunaan metode dan teknik untuk desain dan analisis tes dengan memperbandingkan kemampuan peserta didik pada pembagian kelompok atas, tengah, dan bawah, (3) Konsep reliabilitas skor didefinisikan dari istilah tes paralel, (4) tidak ada dasar teori untuk menentukan bagaimana peserta memperoleh tes yang sesuai dengan kemampuan peserta yang bersangkutan, dan (5) *Standard Error Measurement* (SEM) berlaku pada seluruh peserta tes. Berdasarkan kelemahan-kelemahan tersebut, maka pendekatan Teori Respons Butir muncul untuk mengatasi kelemahan yang ada pada Teori Tes Klasik.

Pendekatan teori respon butir merupakan pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes. Hal ini dikarenakan teori respons butir menggunakan model probabilistik. Model probabilistik bermakna bahwa probabilitas subjek untuk menjawab butir dengan benar bergantung pada kemampuan subjek dan karakteristik butir. Artinya, peserta tes berkemampuan tinggi mempunyai probabilitas menjawab benar lebih besar dibandingkan peserta tes yang berkemampuan rendah.

Teori respon butir memodelkan respon setiap peserta untuk setiap item dalam tes. Item yang mencakup semua jenis butir. Pertanyaan pilihan ganda yang mempunyai jawaban salah dan benar (Qasem, 2013). Teori respons butir perlu menentukan model karakteristik yang digunakan. Model

karakteristik butir dapat berbentuk satu parameter (1PL) atau Model Rasch yang menitik beratkan pada satu parameter butir soal yaitu tingkat kesukaran, dua parameter (2PL) yang melibatkan dua parameter yaitu tingkat kesukaran butir dan daya pembeda, tiga parameter (3PL) yang melibatkan 3 parameter yaitu tingkat kesukaran, daya beda, serta faktor tebakan, atau model lain (Sudaryono, 2012). Dalam penelitian ini akan digunakan model satu parameter atau Model Rasch dengan bantuan program *Winstep versi 3.75.0*.

Pemodelan Rasch bertujuan mengembangkan pengukuran yang objektif. Dalam konteks pemodelan Rasch, antonim dari “pengukuran yang objektif” (*objective measurement*) bukanlah “pengukuran subjektif”, melainkan pengukuran yang hasilnya bergantung pada siapa yang diukur (*test-dependent scoring*). Persentase atau jumlah jawaban benar pada sebuah tes bergantung pada subjek yang diukur (*sample dependent*) yang bersifat deskriptif dan berlaku untuk semua subjek tersebut (Sumintono dan Widhiarso, 2013).

Dr. Rasch pada tahun 1960-an membuat suatu model probabilistik yang didefinisikan sebagai individu yang memiliki tingkat kemampuan abilitas yang lebih besar dibanding individu lainnya seharusnya memiliki peluang yang lebih besar untuk menjawab satu butir soal dengan benar. Dengan prinsip yang sama butir yang lebih sulit menyebabkan peluang individu untuk mampu menjawabnya menjadi kecil (Rasch, 1960 dalam Sumintono dan Widhiarso).

Untuk memastikan pengukuran menjadi objektif maka model pengukuran haruslah memenuhi lima syarat ini: (1) Memberikan ukuran yang linear; (2) Mengatasi data yang hilang; (3) Melakukan proses estimasi yang tepat; (4) Menemukan yang tidak tepat (*misfits*) atau tidak umum (*outliers*); (5) Memberikan instrumen pengukuran yang independen dari parameter yang diteliti. (Sumintono dan Widhiarso, 2013).

Untuk saat ini hanya pemodelan pengukuran Rasch-lah yang memenuhi kelima syarat tersebut. Pengukuran objektif

menghasilkan data yang terbebas dari jenis subjek, karakteristik penilai (*rater*) dan karakteristik alat ukur (Sumintono dan Widhiarso, 2013). Selain itu, berdasarkan pernyataan Ivailo Parchev (2004) yang menyebutkan bahwa untuk butir soal dikotomi dapat menggunakan Model Rasch. Inilah yang menjadi acuan penelitian ini menggunakan pemodelan Rasch untuk menganalisis instrumen tes yang dikembangkan.

Berdasarkan pemaparan diatas penelitian ini dilaksanakan untuk menghasilkan instrumen tes yang handal berupa item tes pilihan ganda pada materi Gaya yang dikembangkan dengan langkah – langkah pengembangan model *Four-D* dan dianalisis menggunakan Model Rasch dengan bantuan program *Winstep versi 3.75.0* yang dapat digunakan untuk mengukur hasil belajar peserta didik dengan baik dan teliti. Dengan demikian diharapkan tidak akan ada peserta didik yang merasa dirugikan atas pengukuran yang kurang teliti dan tidak adil dalam menilai kemampuan mereka dalam belajar.

METODE PENELITIAN

Peneliti menggunakan langkah-langkah *4-D* untuk melakukan penelitian dan pengembangan menurut Mardapi (2012). Langkah-langkah *4-D* meliputi:

Tahap pertama yaitu tahap pendefinisian (*Define*), menetapkan dan mendefinisikan syarat-syarat pembelajaran diawali dengan analisis tujuan dari batasan materi yang dikembangkan. Terdiri dari, (1) Analisis Ujung Depan, (2) Analisis Peserta Didik, (3) Analisis Konsep, (4) Analisis Tugas, (5) Penentuan Tujuan Pembelajaran.

Tahap kedua yaitu tahap perancangan (*Design*), menyiapkan instrumen perangkat pembelajaran. Tahap perencanaan terdiri dari empat langkah yaitu, (1) Menyusun Kisi-Kisi, (2) Menentukan Bentuk Dan Jumlah Soal Yang Sesuai, (3) Menulis Soal, (4) Menentukan Panjang Tes.

Tahap ketiga yaitu tahap pengembangan (*Develop*), menghasilkan perangkat pembelajaran yang sudah direvisi berdasarkan masukan dari pakar. Tahap

pengembangan meliputi, (1) Menelaah Tes, (2) Melakukan Uji Coba Tes, (3) Menganalisis Tes, (4) Memperbaiki Tes, (5) Merakit Tes.

Tahap keempat yaitu tahap penyebaran (*Disseminate*), penggunaan perangkat yang telah dikembangkan pada skala yang lebih luas serta menguji efektivitas penggunaan perangkat di dalam KBM. Tahap penyebaran meliputi (1) Melaksanakan Tes dan (2) Menafsirkan Hasil Tes.

Populasi dalam penelitian ini adalah peserta didik kelas X SMA di Kota Pontianak. Sedangkan sampel pada penelitian ini diambil berdasarkan ukuran sampel dalam pemodelan Rasch dengan nilai $\pm 0,5$ logit serta tingkat kecepecaayaan 99% yaitu kisaran sampel berjumlah 108-243 sampel. Sampel diambil dari rerata nilai Ujian Nasional (UN) IPA tinggi, sedang, dan rendah pada tahun 2017 di Kota Pontianak.

Agar data yang diperoleh memenuhi karakteristik Model Rasch, data tersebut dianalisis menggunakan *software Winstep*. Dalam penelitian ini, analisis data berupa unidimensionalitas, reliabilitas, peta *person-item*, dan tingkat kesukaran butir soal.

HASIL DAN PEMBAHASAN

Hasil

Penelitian ini dilaksanakan di SMA Negeri 3 Pontianak yang berlangsung di kelas X MIPA 1 dan X MIPA 3, SMA Negeri 5 Pontianak yang berlangsung di kelas X MIPA 2 dan X MIPA 4, dan di SMA Negeri 7 Pontianak yang berlangsung kelas X MIPA 1 dan X MIPA 2 pada tahun ajaran 2018/2019.

Penelitian ini dilaksanakan dalam dua tahap, yaitu penelitian dalam skala kecil dan penelitian dalam skala besar. Penelitian skala kecil dilaksanakan di SMA Negeri 8 Pontianak pada tanggal 08 Februari 2019 di kelas X MIPA 2 dengan jumlah peserta didik 34 orang. Penelitian skala besar dilaksanakan di SMA Negeri 3 Pontianak yang berlangsung di kelas X MIPA 1 dengan jumlah peserta didik sebanyak 36 orang dan X MIPA 3 dengan jumlah peserta didik sebanyak 33 orang, SMA Negeri 5 Pontianak

yang berlangsung di kelas X MIPA 2 dengan jumlah peserta didik sebanyak 36 orang dan X MIPA 4 dengan jumlah peserta didik sebanyak 35 orang, dan di SMA Negeri 7 Pontianak yang berlangsung di kelas X MIPA 1 dengan jumlah peserta didik sebanyak 33 orang dan X MIPA 2 dengan jumlah peserta didik sebanyak 29 orang pada tahun ajaran 2018/2019.

Tahap pertama dalam pengembangan tes pada penelitian ini adalah *Define* (Tahap Pendefinisian). Langkah dalam tahap ini yaitu: (1) Analisis ujung depan (*front-end analysis*) Masalah yang diangkat dalam penelitian ini adalah kelayakan instrumen tes yang digunakan untuk mengukur hasil peserta didik di sekolah. (2) Analisis peserta didik. Peneliti menggunakan rerata nilai UN IPA pada mata pelajaran Fisika untuk menentukan kemampuan akademik peserta didik SMA Negeri di kota Pontianak. Kemudian dipilih 3 sekolah dengan kategori tinggi, sedang, dan rendah. (3) Analisis tugas. Analisis yang dilakukan pada tahap ini adalah analisis Kompetensi Inti (KI) dan Kompetensi Dasar (KD) terkait instrumen tes yang akan dikembangkan. (4) Analisis konsep. Konsep yang digunakan untuk mengembangkan instrumen tes berupa materi Gaya dengan sub-materi Gaya berat, Gaya normal, Gaya tegangan tali, Gaya sentripetal dan Hukum Newton tentang gerak, yaitu Hukum I Newton, Hukum II, Newton dan Hukum III Newton. (5) Penentuan tujuan pembelajaran, Menentukan indikator instrumen tes yang harus dicapai oleh peserta didik untuk setiap butir soal berdasarkan KI, KD serta materi yang telah ditentukan.

Tahap kedua adalah *Design* (Tahap Perancangan). Langkah dalam tahap ini yaitu: (1) Penyusunan kisi-kisi dan indikator soal pada materi gerak Gaya. (2) Menentukan bentuk dan jumlah soal yang sesuai. Soal yang dirancang berupa soal pilihan ganda dengan total 40 butir dengan lima pilihan jawaban untuk setiap butir soal. (3) Menulis soal. Tes dirancang dengan jenjang kognitif berdasarkan tingkat taksonomi bloom dari C2-C5 yang telah direvisi menurut Anderson dan Krathwohl. (4) Menentukan panjang tes.

Menentukan waktu yang disediakan berdasarkan tingkat perkembangan peserta didik dan jenjang Pendidikan, yaitu tes dikerjakan selama 2 jam pelajaran atau 2 x 45 menit.

Tahap ketiga adalah *Develop* (Tahap Pengembangan). Langkah dalam tahap ini yaitu: (1) Menelaah tes. Instrumen tes yang telah dirancang kemudian dihitung tingkat keterbacaan menggunakan *microsoft excel*. Hasil perhitungan tingkat keterbacaan instrumen tes yang telah dirancang harus lebih dari atau sama dengan 6. Diperoleh rata-rata tingkat keterbacaan instrumen tes sebesar 7,08. Instrumen tes kemudian divalidasi isi oleh tujuh orang validator yang terdiri dari dua orang dosen fisika dan lima orang guru fisika SMA. Validasi isi bertujuan agar instrumen tes yang dirancang sesuai dengan teori dan materi gerak Gaya di sekolah. Hasil validasi isi dihitung menggunakan rumus Aiken dan didapat rata-rata validitas isi sebesar 0,77 yang menunjukkan instrumen tes yang dikembangkan berada pada kategori tinggi sehingga layak untuk digunakan.

(2) Melakukan coba soal tes atau uji coba skala kecil. Dilakukan di kelas X MIA 2 SMA Negeri 8 Pontianak, uji coba skala kecil bertujuan untuk mempertimbangkan waktu peserta didik untuk mengerjakan tes, alternatif jawaban yang tersedia untuk setiap butir soal tes, dan bahasa yang digunakan dalam penulisan butir tes. (3) Menganalisis butir soal. Dari segi waktu, peserta didik dapat menyelesaikan instrumen tes dalam waktu 70 - 80 menit sehingga 2 jam pelajaran (90 menit) sudah cukup untuk melakukan uji skala besar. Pada alternatif jawaban terdapat beberapa alternatif jawaban yang membingungkan sedangkan pada bahasa yang digunakan sebagian instruksi soal masih membingungkan beberapa peserta didik sehingga perlu diperbaiki. (4) Memperbaiki tes. Perbaikan tes dilakukan dengan menganalisis butir soal yang perlu diperbaiki berdasarkan hasil validitas isi oleh validator serta hasil analisis pada uji coba skala kecil. (5) Merakit tes. Setelah menganalisis dan memperbaiki kesalahan pada butir soal,

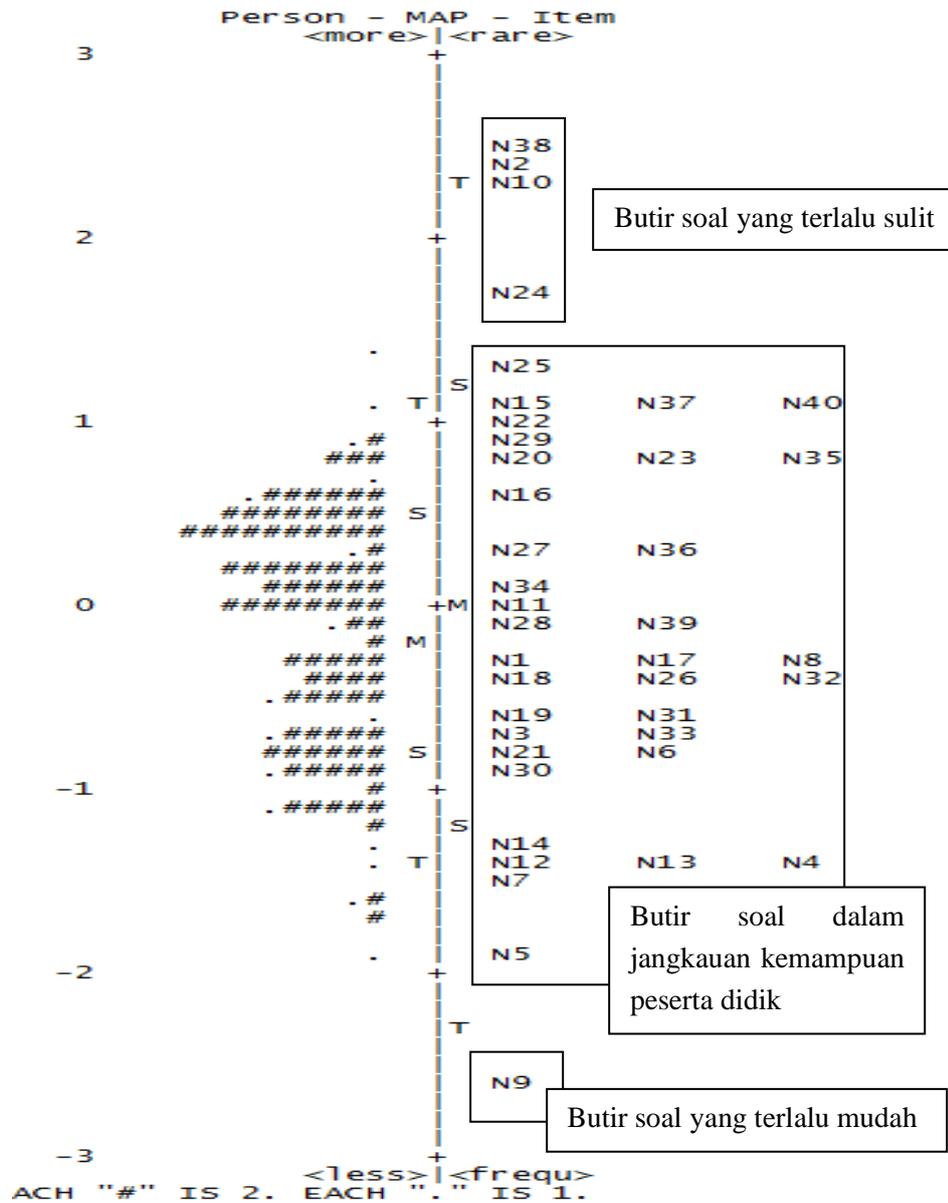
Instrumen tes dirakit kembali sehingga menghasilkan instrumen tes pilihan ganda pada materi Gaya sebanyak 40 butir soal.

Tahap keempat adalah *Desseminate* (Tahap Penyebaran). Langkah dalam tahap ini yaitu: (1) Melaksanakan tes atau uji skala besar. Tes dilaksanakan di kelas X MIPA 1 dan X MIPA 3 SMA Negeri 3 Pontianak, kelas X MIPA 2 dan X MIPA 4 SMA Negeri 5 Pontianak, serta kelas X MIPA 1 dan X MIPA 2 SMA Negeri 7 Pontianak. (2)

Menafsirkan hasil tes. Hasil uji skala besar dianalisis dengan *software* Winstep kemudian diseleksi butir soal yang layak digunakan untuk mengukur hasil peserta didik.

1. Peta *Person – Item*

Soal dikatakan layak apabila soal dapat menjangkau seluruh kemampuan peserta didik dan peserta didik dapat menjangkau seluruh tingkat kesukaran butir soal.



Gambar 1. *Peson-Item Map*

1. Tingkat Kesukaran Butir Soal

Tabel 1. Tingkat Kesukaran Butir Soal

No. Soal	Tingkat Kesukaran	No. Soal	Tingkat Kesukaran
1	-0,27	21	-0,78
2	2,39	22	0,99
3	-0,66	23	0,84
4	-1,38	24	1,67
5	-1,87	25	1,25
6	-0,82	26	-0,42
7	-1,51	27	0,3
8	-0,33	28	-0,13
9	-2,57	29	0,91
10	2,25	30	-0,87
11	0	31	-0,61
12	-1,41	32	-0,36
13	-1,36	33	-0,75
14	-1,33	34	0,08
15	1,1	35	0,76
16	0,61	36	0,26
17	-0,31	37	1,07
18	-0,41	38	2,52
19	-0,58	39	-0,13
20	0,76	40	1,08

2. Unidimensionalitas

Unidimensionalitas menyatakan apakah instrumen tes yang dikembangkan mampu mengukur apa yang seharusnya diukur. Hasil analisis unidimensionalitas yang diperoleh dari uji coba skala besar sebesar 26,2%, dan termasuk ke dalam kategori cukup.

dapat dipercaya apabila memberikan hasil pengukuran yang relatif tetap. Reliabilitas instrumen tes yang dikembangkan dalam penelitian ini berdasarkan *alpha cronbach* didapat nilai sebesar 0,68 termasuk ke dalam kategori cukup (0,6-0,7 = cukup).

3. Reliabilitas

Menurut Azwar 2013, reliabilitas artinya dapat dipercaya. Suatu alat ukur dikatakan

Tabel 2. Hasil Analisis Butir Soal

Nomor Soal	Karakteristi Tes yang Dikembangkan					Keterangan
	1	2	3	4	5	
1	√	26,2%	0,68	-0,27	7,91	Dipakai
2	-	26,2%	0,68	2,39	6,92	Dibuang
3	√	26,2%	0,68	-0,66	7,27	Dipakai
4	√	26,2%	0,68	-1,38	6,92	Dibuang

Nomor Soal	Karakteristi Tes yang Dikembangkan					Keterangan
	1	2	3	4	5	
5	√	26,2%	0,68	-1,87	9,72	Dibuang
6	√	26,2%	0,68	-0,82	8,26	Dipakai
7	√	26,2%	0,68	-1,51	6,15	Dibuang
8	√	26,2%	0,68	-0,33	8,06	Dipakai
9	-	26,2%	0,68	-2,57	6,43	Dibuang
10	-	26,2%	0,68	2,25	7,30	Dibuang
11	√	26,2%	0,68	0	6,07	Dipakai
12	√	26,2%	0,68	-1,41	6,79	Dibuang
13	√	26,2%	0,68	-1,36	6,07	Dibuang
14	√	26,2%	0,68	-1,33	9,00	Dibuang
15	√	26,2%	0,68	1,1	6,06	Dibuang
16	√	26,2%	0,68	0,61	6,49	Dipakai
17	√	26,2%	0,68	-0,31	7,69	Dipakai
18	√	26,2%	0,68	-0,41	6,36	Dipakai
19	√	26,2%	0,68	-0,58	8,76	Dipakai
20	√	26,2%	0,68	0,76	6,59	Dipakai
21	√	26,2%	0,68	-0,78	7,36	Dipakai
22	√	26,2%	0,68	0,99	6,71	Dipakai
23	√	26,2%	0,68	0,84	6,13	Dipakai
24	-	26,2%	0,68	1,67	6,33	Dibuang
25	√	26,2%	0,68	1,25	6,23	Dibuang
26	√	26,2%	0,68	-0,42	7,59	Dipakai
27	√	26,2%	0,68	0,3	6,27	Dipakai
28	√	26,2%	0,68	-0,13	6,85	Dipakai
29	√	26,2%	0,68	0,91	7,75	Dipakai
30	√	26,2%	0,68	-0,87	6,12	Dipakai
31	√	26,2%	0,68	-0,61	6,70	Dipakai
32	√	26,2%	0,68	-0,36	9,10	Dipakai
33	√	26,2%	0,68	-0,75	7,58	Dipakai
34	√	26,2%	0,68	0,08	6,96	Dipakai
35	√	26,2%	0,68	0,76	7,43	Dipakai
36	√	26,2%	0,68	0,26	6,07	Dipakai
37	√	26,2%	0,68	1,07	6,09	Dibuang
38	-	26,2%	0,68	2,52	7,56	Dibuang
39	√	26,2%	0,68	-0,13	6,85	Dipakai
40	√	26,2%	0,68	1,08	7,20	Dibuang

Keterangan:

1 = Peta peson-item (√ = Terpenuhi, - = Tidak terpenuhi)

2 = Unidimensionalitas (> 20 %)

3 = Reliabilitas (Minimal 0,60)

4 = Tingkat kesukaran (Rentang -1 sampai 1)

5 = Tingkat keterbacaan (Minimal RI ≥ 6)

Pembahasan

Peta item-person adalah peta variabel yang menampilkan sebaran kemampuan peserta tes dan tingkat kesukaran butir soal

pada skala yang sama. Dari peta variabel ini diperoleh informasi sejauh mana butir – butir soal penyusun tes yang dikembangkan layak

digunakan untuk mengukur kemampuan peserta didik. Dari peta variabel diperoleh tiga jenis kelompok butir soal, yaitu kelompok butir soal yang tidak dapat dijangkau oleh peserta didik dengan kemampuan tertinggi dengan nilai $logit > 2$, kelompok butir soal yang dapat dijangkau oleh seluruh kemampuan peserta didik yang terletak pada nilai $logit -2$ hingga 2 , dan kelompok butir soal yang memiliki tingkat kesukaran kurang dari kemampuan peserta didik terendah dengan nilai $logit < -2$.

1. Kelompok butir soal yang tidak dapat dijangkau oleh peserta didik dengan kemampuan tertinggi nilai $logit > 2$ terdapat 4 butir soal. Butir – butir soal ini memiliki tingkat kesukaran paling sulit untuk dikerjakan peserta didik dan tidak dapat dijangkau oleh peserta didik dengan kemampuan tertinggi sekalipun didalam tes. Empat butir soal tersebut yaitu soal nomor 24, 10, 2, dan 38. Butir soal seperti ini kurang memiliki fungsi tidak layak digunakan untuk mengukur kemampuan peserta didik karena tingkat kesukaran yang sangat tinggi dapat menyebabkan soal tes menjadi tidak valid. Hal ini bisa disebabkan karena soal yang dikembangkan tidak sesuai dengan materi pembelajaran yang diterima oleh peserta didik (Sukardi, 2012).

2. Kelompok butir soal yang dapat dijangkau oleh seluruh kemampuan peserta didik terdapat 25 butir soal nilai $logit -2$ sampai 2 . Butir – butir soal ini memiliki tingkat kesukaran yang bervariasi dari mudah hingga sulit dan tersebar merata dalam skala yang sama dengan seluruh kemampuan peserta didik. Dua puluh lima butir soal tersebut yaitu soal nomor 5, 7, 12, 13, 4, 14, 30, 21, 6, 3, 33, 19, 31, 18, 26, 32, 1, 17, 8, 28, 39, 11, 34, 27, 36, 16, 20, 23, 35, 29, 22, 15, 37, 40, dan 25. Butir soal seperti ini layak digunakan untuk mengukur kemampuan peserta didik. Karena memiliki tingkat kesukaran yang beragam dari yang paling mudah, sedang, dan sulit.

3. Kelompok butir soal yang memiliki tingkat kesukaran kurang dari kemampuan peserta didik terendah terdapat satu butir soal nilai $logit > -2$, yaitu soal nomor 9. Butir soal

ini memiliki tingkat kesukaran paling mudah. Butir soal yang terlalu mudah tidak dapat merangsang proses berfikir peserta didik, selain itu butir soal ini tidak memiliki daya diskriminasi sehingga tidak dapat membedakan kemampuan peserta didik yang satu dengan yang lainnya.

Jika dilihat secara keseluruhan dapat diketahui bahwa butir dalam tes dapat menjangkau semua interval dalam peta butir. Hal ini menunjukkan instrumen tes yang dikembangkan layak digunakan untuk mengukur kemampuan peserta didik.

Unidimensionalitas adalah ukuran yang penting untuk mengevaluasi apakah instrumen tes pilihan ganda yang dikembangkan mampu mengukur apa yang seharusnya diukur. Analisis model Rasch menggunakan analisis komponen utama (*Partial Component Analysis*) dari residual, yaitu mengukur sejauh mana keragaman dari instrumen tes yang dikembangkan mengukur apa yang seharusnya diukur. Dari hasil analisis ini diperoleh hasil pengukuran *Raw variance* sebesar 26,2%. Hal ini menunjukkan bahwa persyaratan unidimensionalitas minimal sebesar 20% dapat dipenuhi. Dengan demikian instrumen tes pilihan ganda yang dikembangkan dalam penelitian ini cukup valid untuk mengukur kemampuan peserta didik.

Reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya (Azwar, 2013). Reliabilitas instrumen tes dalam penelitian ini dianalisis menggunakan program analisis butir soal Winstep menggunakan model Rasch. Dari hasil analisis program Winstep diperoleh nilai *alpha cronbach* sebesar 0,68 termasuk ke dalam kategori cukup yaitu dengan nilai dikisaran antara 0,6 s.d 0,7, reliabilitas peserta didik atau *person reliability* sebesar 0,65 termasuk ke dalam kategori lemah karena memiliki nilai kurang dari 0,67, dan reliabilitas item atau *item reliability* sebesar 0,98 termasuk ke dalam kategori istimewa karena memiliki nilai lebih dari 0,94.

Reliabilitas peserta didik sangat rendah dengan kategori lemah, hal ini menunjukkan bahwa tingkat konsistensi peserta didik

dalam mengerjakan tes lemah. Sedangkan reliabilitas item sangat tinggi dengan kategori istimewa ini menunjukkan bahwa item tes yang dikembangkan memiliki tingkat konsistensi yang tinggi dalam pengukuran kemampuan peserta didik.

Menurut Hambleton & Swaminathan dalam (Istiyono, 2017), tingkat kesukaran untuk tes yang baik adalah tes yang memiliki tingkat kesukaran yang bervariasi. Tingkat kesukaran dalam tes pilihan ganda yang dikembangkan memiliki kriteria tes dengan kategori tingkat kesukaran mudah, sedang, dan sukar. Namun ada beberapa butir soal yang memiliki tingkat kesukaran paling tinggi dan tingkat kesukaran paling rendah yang berjumlah 15 butir soal. Soal yang terlalu sulit atau terlalu mudah tidak dapat menjalankan fungsinya sebagai alat ukur yang baik (Sumintono dan Widhiarso, 2013). Karena soal yang dikembangkan dalam penelitian ini untuk fungsi formatif maka butir soal yang terlalu sulit atau terlalu mudah akan dibuang. Selain itu, jika digunakan untuk tes final maka butir – butir soal tersebut akan menurunkan motivasi peserta didik.

Dengan demikian berdasarkan hasil analisis tingkat kesukaran butir soal, soal yang layak digunakan berjumlah 25 butir soal yaitu soal nomor 1, 3, 6, 8, 17, 18, 19, 21, 26, 28, 30, 31, 32, 33, 39, 11, 16, 20, 22, 23, 27, 29, 34, 35, dan 36.

Tingkat keterbacaan dinyatakan dalam *readability index* yang mencakup variabel panjang kalimat (*sentence length*) dan panjang kata (*word length*). Tingkat keterbacaan instrumen tes pilihan ganda menunjukkan nilai rata-rata sebesar 7,08. Karena nilai tingkat keterbacaan menunjukkan nilai > 6, maka tes ini sesuai untuk digunakan pada jenjang SMA (Sutrisno dalam Mahmuda, 2011).

Berdasarkan hasil analisis diatas, dapat disimpulkan instrumen tes yang dikembangkan layak digunakan untuk mengukur kemampuan peserta didik dengan baik. Namun, masih terdapat beberapa butir soal yang terlalu sulit dan terlalu mudah. Sehingga perlu dilakukan analisa lebih lanjut

untuk memilih butir soal yang layak digunakan untuk mengukur tingkat hasil belajar peserta didik.

Selain itu karena dalam penelitian ini banyak peserta didik yang tidak mengisi lembar jawaban pada beberapa nomor soal dalam tes maka nilai *Infit* MNSQ perlu dianalisis dan dilaporkan. Salah satu statistik fit item yang utama adalah *infit mean square* (INFIT MNSQ). *Infit mean square* mengukur konsistensi kecocokan peserta didik dengan kurva karakteristik item untuk setiap item dengan pertimbangan yang diberikan kepada person yang dekat dengan tingkat probabilitas 0,5 (Alagumalai. dkk, 2005).

Langkah pertama yaitu menganalisis nilai MNSQ untuk mengevaluasi kecocokan, selama nilai MNSQ berada dalam rentang kesesuaian yang diterima, nilai ZSTD diabaikan (Boone. dkk, 2014). Berdasarkan tabel tingkat kesukaran pada gambar 4.3 terdapat 3 butir soal yang memiliki nilai *infit* MNSQ dan *outfit* MNSQ diluar rentang nilai yang diterima (memiliki nilai diatas 1,5), yaitu soal nomor 24, 10 dan 38 dengan nilai berturut – turut 1,51; 1,71; dan 2,62. Maka nilai ZSTD juga perlu dianalisis.

Langkah kedua menganalisis ZSTD, analisis ini dilakukan dengan melihat nilai *infit* ZSTD dan *outfit* ZSTD. Nilai yang diterima memiliki rentang antara -1,9 s.d 1,9 (Sumintono dan Widhiarso, 2013). Jika diluar dari rentang tersebut maka soal tersebut dibuang atau diganti dengan soal yang baru. Berdasarkan Berdasarkan tabel tingkat kesukaran pada gambar 4.3 terlihat bahwa soal nomor 24, 10 dan 38 memiliki nilai diatas 1,9 yaitu dengan nilai berturut – turut 2,6; 2,5; dan 4,1. Dengan demikian 3 soal tersebut tidak dapat diterima dan dibuang atau diganti dengan soal yang baru.

Berdasarkan hasil analisis butir soal instrumen tes pilihan ganda pada materi Gaya yang dikembangkan menggunakan model Rasch diperoleh 25 butir soal yang layak digunakan untuk mengukur kemampuan peserta didik. Butir – butir soal tersebut yaitu butir soal nomor 1, 3, 6, 8, 11, 16, 17, 18, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, dan 39.

SIMPULAN DAN SARAN

Simpulan

Tes pilihan ganda yang dikembangkan pada materi Gaya untuk SMA layak digunakan sebagai alat ukur untuk mengukur hasil belajar peserta didik dan menghasilkan tes pada materi Gaya. Hasil penelitian ini dapat disimpulkan sebagai berikut:

Langkah penyusunan tes terdiri dari empat tahap, yaitu tahap pendefinisian (analisis ujung depan, analisis peserta didik, analisis konsep, analisis tujuan pembelajaran), tahap perancangan (menyusun kisi – kisi, menentukan bentuk dan jumlah soal yang sesuai, menulis soal, menentukan panjang tes, rancangan awal), tahap pengembangan, menelaah tes, melakukan uji coba skala kecil, menganalisis butir tes, memperbaiki tes, merakit tes), tahap penyebaran (pelaksanaan tes skala besar, menganalisis butir tes dan menafsirkan tes).

Pengembangan tes pilihan ganda pada materi Gaya telah memenuhi karakteristik tes yang baik, yaitu 1) butir soal tersebar merata pada peta person-item dan dapat menjangkau seluruh kemampuan akademik peserta didik, 2) unidimensionalitas cukup dengan *Raw variance* sebesar 26,2%, 3) reliabilitas cukup, yaitu sebesar 0,68, 4) memiliki tingkat kesukaran mudah sampai sukar yaitu sebanyak 25 butir soal, dengan rentang *logit* -1 sampai 1, 5) tingkat keterbacaan tes sesuai dengan jenjang yang di ukur, yaitu 7,08.

Saran

1. Untuk penelitian selanjutnya yang serupa sebaiknya untuk menghasilkan instrumen tes yang benar – benar layak digunakan untuk mengukur hasil belajar peserta didik, uji coba instrumen tes yang dikembangkan dilakukan sebanyak 3 kali dan dengan jumlah subjek lebih dari 200 orang peserta didik.
2. Sebelum melaksanakan uji coba instrumen tes yang dikembangkan terlebih dahulu berkoordinasi dengan guru mata pelajaran Fisika untuk bisa mendampingi dalam pelaksanaan uji coba instrumen tes di kelas agar peserta didik lebih serius dalam mengerjakan tes.

3. Pelaksanaan uji coba instrumen tes yang dikembangkan akan lebih baik jika disertai angket respon peserta didik agar diperoleh data tambahan mengenai kualitas instrumen tes yang dikembangkan.

DAFTAR RUJUKAN

- Alagumalai, S., Curtis, D. D., & Hungi, N. (2005). *Applied Rasch Measurement: A Book of Exemplars*. Dordrecht: Springer.
- Arifin, Z. (2009). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Azwar, S. (2012). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Depdiknas. (2004). *Kurikulum Kerangka Dasar 2004*. Jakarta. Depdiknas.
- Hambleton, R.K. & H. Swaminathan. (1985). *Fundamentals of item Response Theory*. London: Sage Publications.
- Istiyono, E. (2014). *Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Peserta Didik SMA. Jurnal Penelitian dan Evaluasi Pendidikan, 14, 1-12*.
- Mahmudah, D. (2011). *Secondary Analysis Tentang Tes Skripsi-Skripsi Mahasiswa Pendidikan Fisika FKIP Untan Tahun 2007-2009 Pada Materi Mekanika*. Pontianak: FKIP Untan (Skripsi).
- Naga, D. S. (1992). *Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Purwanto. (2008). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar.
- Sudaryono. (2012). *Dasar-Dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu.
- Sudijono, A. (2012). *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Pers.
- Sukardi. 2012. *Evaluasi Pendidikan Prinsip & Operasionalnya*. Yogyakarta: Bumi Aksara.
- Sumintono, B. dan Widhiarso, W. (2013). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. Cimahi: Trim Komunikata.

