

## Artikel Penelitian/Article Review

# Pengembangan Instrumen Kognitif Fisika Siswa SMP

<sup>1</sup>Mustika Wati, <sup>2</sup>Saiyidah Mahtari

<sup>1,2</sup>Program Studi Pendidikan Fisika, FKIP Universitas Lambung Mangkurat, Jl. Brigjen H. Hasan Basri No.87, Kalimantan Selatan 70123, Indonesia

Email: mustika\_pfis@unlam.ac.id

### ARTICLE INFO

#### Article history

Received: March 2017

Revised: April 2017

Accepted: May 2017

Published: June 2017

#### Keywords

Development;  
cognitive instrument;  
rasch model

### ABSTRACT

**[Title: Development of Physics Cognitive Instruments Of Junior High School Students]** This study aims to develop cognitive instruments on energy and business materials, simple aircraft and light quality. The specific purpose of this study is to describe: (1) the validity of cognitive instruments seen from Outfit MNSQ, ZSTD, and Pt Measure Corr; (2) the reliability of cognitive instruments seen from the values of individual separation, grain separation, and Cronbach Alpha; (3) the degree of difficulty seen from the logit value in the measurement column; This development research refers to the development model of Borg and Gall with the subject of research of 200 students of class VIII originating from several SMPN in Banjarmasin. Instrument collecting data in the form of validation and cognitive instruments. Data analysis techniques are descriptive and quantitative. The results showed: (1) five items of description on light material, nine items about the description of simple plane materials and nine items about the description of energy materials and businesses that are categorized as valid; (2) reliability of cognitive instruments seen from individual separation values, grain separation values, and Cronbach Alpha values are categorized as very reliable; (3) the difficulty level of questions seen from the logit of categorized is very difficult, difficult, easy, and very easy.

### INFO ARTIKEL

#### Sejarah Artikel

Dikirim: Maret 2017

Direvisi: April 2017

Diterima: Mei 2017

Dipublikasi: Juni 2017

#### Kata kunci

Pengembangan;  
Instrumen kognitif;  
rasch model

### ABSTRAK

Penelitian ini bertujuan untuk mengembangkan instrument kognitif pada materi energi dan usaha, pesawat sederhana dan cahaya yang berkualitas. Tujuan khusus penelitian ini adalah mendeskripsikan: (1) validitas instrumen kognitif dilihat dari *Outfit MNSQ*, *ZSTD*, dan *Pt Measure Corr*; (2) reliabilitas instrumen kognitif yang dilihat dari nilai separasi individu, separasi butir, dan *Alpha Cronbach*; (3) tingkat kesukaran yang dilihat dari nilai *logit* pada kolom pengukuran. Penelitian pengembangan ini mengacu pada model pengembangan Borg and Gall dengan subjek penelitian 200 siswa kelas VIII yang berasal dari beberapa SMPN di Banjarmasin. Instrumen pengumpul data berupa instrumen validasi dan kognitif. Teknik analisis data berupa deskriptif dan kuantitatif. Hasil penelitian menunjukkan: (1) lima butir soal uraian pada materi cahaya, sembilan butir soal uraian materi pesawat sederhana dan sembilan butir soal uraian materi energi dan usaha yang terkategori valid; (2) reliabilitas instrumen kognitif dilihat dari nilai separasi individu, nilai separasi butir, serta nilai *Alpha Cronbach* yang masuk kategori sangat reliabel; (3) tingkat kesukaran soal yang dilihat dari nilai *logit* terkategori sangat sulit, sulit, mudah, dan sangat mudah.

### How to Cite this Article?

Wati, M., & Mahtari, S. (2017). Pengembangan Instrumen Kognitif Fisika Siswa SMP. *Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika*, 1(1), 45-56.

## PENDAHULUAN

Salah satu kompetensi inti guru Peraturan Menteri Pendidikan Nasional Nomor 16 Tahun 2007 menyelenggarakan penilaian hasil belajar menyatakan bahwa. Sehingga kompetensi yang harus dimiliki guru adalah mengembangkan instrumen penilaian, evaluasi proses, dan hasil belajar. Kebijakan dari pemerintah yang memberikan kewenangan pada guru untuk membuat soal Ujian Sekolah Berstandar Nasional (USBN) menunjukkan bahwa kemampuan guru dalam membuat soal sangat diperlukan. Pada rambu-rambu USBN dijelaskan bahwa jumlah butir soal 20% - 25% disiapkan oleh pemerintah dan 75% - 80% disiapkan oleh MGMP untuk SMP/MTs dan SMA/MA di bawah koordinasi Dinas Pendidikan. Dengan wewenang 75% - 80% guru diharapkan dapat membuat kisi-kisi/indikator soal, penulisan soal, dan penskoran soal baik pilihan ganda maupun uraian beserta rubrik penilaiannya. Basuki dan Hariyanto (2015) menyatakan bahwa ciri - ciri pokok suatu soal dikatakan baik diantaranya dapat dipercaya atau valid, objektif, dan praktis.

Berdasarkan hasil wawancara dengan guru-guru di beberapa SMPN di Banjarmasin, instrumen penilaian yang dibuat belum dilakukan analisis butir soal. Artinya dapat dikatakan soal yang diberikan kepada peserta didik belum dapat dinyatakan sah dan objektif. Penilaian yang sah dan objektif dapat diperoleh dengan menggunakan tes yang memuat soal atau pertanyaan-pertanyaan yang dapat mencerminkan kemampuan yang diukur (Mustari, 2016). Sehingga kemampuan siswa belum dapat terukur dengan baik melalui soal-soal yang belum dilakukan analisis butir soal.

Mengatasi permasalahan mengenai penilaian tersebut maka salah satu caranya yaitu membuat instrumen penilaian sesuai kompetensi yang harus dicapai dan sudah diuji validitas dan reliabilitas. Instrumen penilaian yang dibuat berupa soal uraian. Selama ini kebanyakan guru fisika di sekolah menggunakan soal pilihan ganda dalam tes yang mereka buat. Guru harus lebih sering memberikan tes esai. Dirman dan Juarsih (2014) menyebutkan bahwa tes esai dapat mengukur kegiatan-kegiatan belajar yang sulit diukur oleh tes obyektif. Tes esai menuntut anak untuk menguraikan jawabannya dengan kata-kata sendiri dan cara tersendiri. Agar guru dapat mengetahui sampai mana pemahaman siswa pada pelajaran yang telah diberikan, sehingga perlu dibuat soal esai yang menekankan pada proses siswa mendapatkan jawaban tersebut.

Terdapat keunggulan-keunggulan tes uraian seperti yang dijelaskan Rasyid dan Mansyur (2011) antara lain adalah dapat mengukur proses mental yang tinggi atau aspek kognitif tingkat tinggi; dapat mengembangkan kemampuan berbahasa, baik lisan maupun tulisan, dengan baik dan benar dengan kaidah-kaidah bahasa; dapat melatih kemampuan berfikir teratur atau penalaran, yakni berpikir logis, analitis, dan sistematis; mengembangkan keterampilan pemecahan masalah (*problem solving*); adanya keuntungan teknis seperti mudah membuat soalnya sehingga tanpa memakan waktu yang lama, guru dapat secara langsung melihat proses berpikir siswa.

Basuki dan Hariyanto (2015) mengemukakan ciri-ciri pokok suatu tes dikatakan baik antara lain, sah atau valid, dapat dipercaya, objektif, serta praktis. Validitas suatu instrumen dalam Sukardi (2015) adalah derajat yang menunjukkan

kemampuan suatu tes mengukur apa yang seharusnya diukur. Adapun Purwanto (2013) menyatakan bahwa validitas (kesahihan) merupakan kualitas menunjukkan hubungan antara suatu pengukuran (diagnosis) dengan arti atau tujuan kriteria belajar atau tingkah laku. Validitas merupakan syarat terpenting dalam suatu alat evaluasi. Sehingga dapat dikatakan bahwa validitas adalah suatu kualitas yang menunjukkan fakta empiris dan alasan teoritis tentang apa yang hendak diukur.

Menurut Purwanto (2013) keandalan (reliabilitas) adalah ketetapan atau ketelitian suatu alat evaluasi. Sukardi (2015) mendefinisikan reliabilitas sebagai konsistensi atau keajegan. Suatu instrumen evaluasi, dikatakan mempunyai nilai reliabilitas yang tinggi, apabila tes yang dibuat mempunyai hasil yang konsisten dalam mengukur yang hendak diukur.

Menurut Arikunto (2013) soal dikatakan baik jika tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah tidak merangsang siswa berusaha keras memecahkannya, sebaliknya soal yang terlalu sukar akan mematahkan semangat siswa dan putus asa untuk mencoba lagi. Rasyid dan Mansyur (2011) mendefinisikan tingkat kesukaran suatu butir soal sebagai proporsi atau presentase subjek yang menjawab butir tes tertentu dengan benar.

Ada delapan langkah yang perlu ditempuh dalam mengembangkan tes hasil atau prestasi belajar, yaitu:

(1) Menyusun spesifikasi tes

Menyusun spesifikasi tes yaitu berisi tentang uraian yang menunjukkan keseluruhan karakteristik yang harus dimiliki suatu tes. Penyusun spesifikasi tes mencakup: (a) menentukan tujuan tes, (b) menyusun kisi-kisi tes, (c) memilih bentuk tes, (d) menentukan panjang tes.

(2) Menulis soal tes

Penulisan soal tes merupakan langkah menjabarkan indikator menjadi pertanyaan-pertanyaan yang karakteristiknya sesuai dengan perincian kisi-kisi yang telah dibuat.

(3) Menelaah soal tes

Hal ini diperlukan untuk memperbaiki soal jika ternyata dalam pembuatannya masih ditemukan kekurangan atau kesalahan.

(4) Melakukan ujicoba tes

Sebelum soal digunakan dalam tes yang sesungguhnya, uji coba perlu dilakukan untuk semakin memperbaiki kualitas soal.

(5) Menganalisis butir soal

Melalui analisis butir soal dapat diketahui antara lain: tingkat kesukaran butir soal, daya pembeda, dan juga efektivitas pengecoh.

(6) Memperbaiki tes

Langkah ini dilakukan atas butir soal, yaitu memperbaiki masing-masing butir soal yang ternyata masih belum baik.

(7) Merakit tes

Keseluruhan butir perlu disusun secara hati-hati menjadi kesatuan soal tes yang terpadu, dalam merakit soal, hal-hal yang dapat memengaruhi validitas soal seperti nomor urut soal, pengelompokan bentuk soal, *lay out* dan sebagainya harus diperhatikan.

(8) Melaksanakan Tes

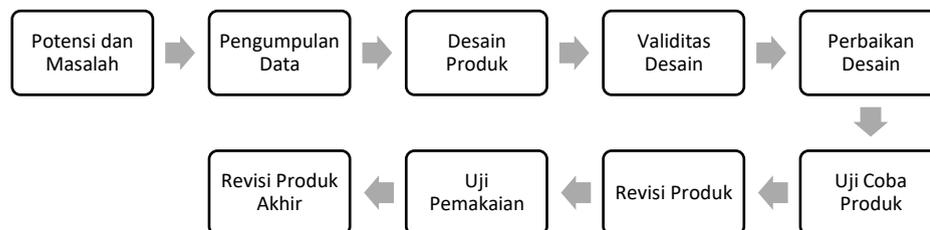
Setelah langkah menyusun tes selesai, langkah selanjutnya adalah melaksanakan tes. Tes yang telah disusun diberikan kepada testee untuk diselesaikan.

#### (9) Menafsir Hasil Tes

Hasil tes menghasilkan data kuantitatif yang berupa skor. Skor ini kemudian ditafsirkan sehingga menjadi nilai, yaitu rendah, menengah, atau tinggi. Tinggi rendahnya nilai ini selalu dikaitkan dengan acuan penilaian yang terdiri dari acuan penilaian norma dan kriteria. Tujuan umum penelitian ini adalah mendeskripsikan kualitas instrumen kognitif fisika yang dikembangkan dilihat dari validitas, reliabilitas dan tingkat kesukaran.

### METODE

Penelitian ini mengembangkan instrumen penilaian kognitif berupa tes atau butir-butir soal. Penelitian ini merupakan penelitian *Research and Development* (R&D) dengan menggunakan model prosedur dari Borg and Gall yang diadopsi oleh Sugiono (2015).



Gambar 1 Langkah-langkah penggunaan Metode R &D  
Sumber: Adaptasi Sugiyono, 2015

Pada tahap potensi dan masalah, peneliti melakukan analisis kebutuhan untuk mengetahui instrumen penilaian yang digunakan oleh guru SMP Negeri di Kota Banjarmasin dalam pembelajaran IPA. Pada tahap pengumpulan data ini dilakukan dengan wawancara terhadap guru IPA Kelas VIII di SMP Banjarmasin. Informasi yang didapatkan dari kegiatan wawancara ialah terkait instrumen penilaian yang digunakan oleh guru IPA dalam penilaian.

Pada tahap desain produk, peneliti mendesain produk yang akan dikembangkan yaitu pembuatan instrumen kognitif berupa soal uraian pada materi Cahaya, usaha energi dan pesawat sederhana untuk siswa SMP kelas VIII di Kota Banjarmasin. Kemudian tahap validasi desain, tahap ini merupakan proses penilaian rancangan produk yang dilakukan dengan memberi penilaian berdasarkan pemikiran rasional tanpa uji coba lapangan kepada dua validator ahli yang sudah ditentukan. Analisis validasi desain dilakukan berdasarkan uji pakar dengan memberikan lembar validasi instrumen kepada dua orang validator yang telah ditentukan dengan kriteria pada Tabel 1 berikut.

Tabel 1. Kriteria validitas instrumen

Nilai Rata-rata Validator	Kategori	Keterangan
>4,2	Sangat Valid	Dapat digunakan tanpa revisi
>3,4 - 4,2	Valid	Digunakan namun perlu revisi kecil
>2,6 - 3,4	Cukup Valid	Disarankan tidak perlu digunakan karena perlu revisi besar
>1,8 - 2,6	Kurang Valid	Tidak boleh digunakan, perlu revisi besar-besarran
≤1,8	Tidak Valid	Tidak boleh dipergunakan

(Widoyoko, 2013)

Tahap revisi desain dilakukan untuk memperbaiki produk yang telah dibuat sebelum produk tersebut diujicobakan. Pada tahap ini peneliti memperbaiki kembali desain produk yang telah divalidasi sesuai saran-saran yang diberikan oleh kedua validator. Pada tahap ujicoba produk, produk yang telah dibuat diujicobakan pada siswa kelas IX di SMP dalam skala kecil yang telah mempelajari materi tersebut. Analisis validitas soal dilakukan dengan menggunakan rumus Pearson. Reliabilitas soal dalam ujicoba produk dilakukan dengan persamaan *Alpha Cronbach*.

Revisi produk dilakukan untuk menyempurnakan kembali produk yang telah dikembangkan dan disesuaikan dengan kondisi nyata di lapangan berdasarkan hasil uji coba produk yang telah dilakukan. Pada tahap ujicoba pemakaian, produk yang dihasilkan diujicobakan ke siswa kelas VIII SMP di kota Banjarmasin seluruhnya berjumlah 200 siswa yang berasal dari sekolah SMPN 22 Banjarmasin, SMPN 23 Banjarmasin, SMPN 28 Banjarmasin, SMPN 27 Banjarmasin, SMP Negeri 14 Banjarmasin, SMP Negeri 4 Banjarmasin, dan SMP Negeri 20 Banjarmasin.

Analisis validitas soal dengan menggunakan program *Rasch* dinyatakan dengan nilai MNSQ, nilai ZSTD, dan nilai korelasi pengukuran. Adapun kriteria untuk menginterpretasikan nilai MNSQ dapat dilihat dengan menggunakan Tabel 2 berikut.

Tabel 2. Interpretasi Nilai MNSQ

Nilai MNSQ	Implikasi pada Pengukuran
> 2,0	Menurunkan kualitas sistem pengukuran.
1,5 - 2,0	Kurang bagus untuk pembuatan instrumen tapi tidak menurunkan kualitas.
0,5-1,5	Kondisi yang baik untuk pengukuran.
< 0,5	Kurang produktif untuk pengukuran, namun tidak menurunkan kualitas; kemungkinan bisa menyebabkan kesalahan dengan reliabilitas yang tinggi.

(Sumber: Sumintono dan Widhiarso, 2015)

Adapun kriteria untuk menginterpretasikan nilai ZSTD.

Tabel 3. Interpretasi Nilai ZSTD

Nilai ZSTD	Implikasi pada Pengukuran
$\geq 3,0$	Data tidak diharapkan jika sesuai dengan model (secara sempurna). Namun dengan sampel yang besar, ketidaksesuaiannya mungkin lebih kecil.
2,0 - 2,9	Data tampak tidak dapat diprediksi.
-1,9 - 2,9	Data mempunyai perkiraan yang logis.
$\leq -2$	Data terlalu mudah diprediksi.

(Sumber: Sumintono dan Widhiarso, 2015)

Sedangkan kriteria untuk korelasi (Pt. Mean Corr) terletak antara 0,4 sampai 0,85 (Sumintono dan Widhiarso, 2015).

Menurut Sumintono dan Widhiarso (2015) nilai reliabilitas dalam pemodelan Rasch ditunjukkan dengan nilai separasi individu dan separasi butir. Separasi individu menunjukkan seberapa baik butir soal dalam tes menyebar sepanjang rentang atau kontinum kemampuan *logit*. Semakin besar harga separasi individu, semakin baik tes yang disusun karena butir-butir di dalamnya mampu menjangkau individu dengan kemampuan di tingkat tinggi sampai ke yang rendah. Separasi butir menunjukkan seberapa besar sampel yang dikenakan pengukuran tersebar sepanjang skala interval linier. Semakin tinggi nilai separasi butir, semakin baik pengukuran yang dilakukan.

Adapun kriteria untuk menginterpretasikan nilai *person reliability* dan *item reliability* sebagai berikut.

Tabel 4 Interpretasi nilai *person reliability* dan *item reliability*

Nilai <i>person reliability</i> dan <i>item reliability</i>	Kriteria
$> 0,94$	Istimewa
0,91 - 0,94	Bagus sekali
0,81 - 0,90	Bagus
0,67 - 0,80	Cukup
$< 0,67$	Lemah

(Sumintono dan Widhiarso, 2015)

Selain itu, nilai Alpha Cronbach juga digunakan untuk mengukur interaksi antara individu dengan butir-butir keseluruhan dapat diinterpretasikan dengan menggunakan Tabel 5 berikut.

Tabel 5. Interpretasi Nilai *Alpha Cronbach*

Nilai <i>Alpha Cronbach</i>	Kriteria
$< 0,5$	Buruk
0,5 - 0,6	Jelek
0,6 - 0,7	Cukup
0,7 - 0,8	Bagus
$> 0,8$	Bagus sekali

(Sumintono dan Widhiarso, 2015)

Pada pemodelan Rasch, Sumintono dan Widhiarso (2015) menjelaskan tingkat kesukaran butir soal dapat dilihat dari nilai *logit* yang terdapat pada kolom pengukuran soal. Nilai *logit* yang tinggi menunjukkan tingkat kesulitan soal yang tinggi. Pada kolom pengukuran soal juga terdapat informasi nilai deviasi standar. Jika nilai dikombinasikan dengan rata-rata nilai *logit*, tingkat kesulitan butir dapat dikelompokkan.

Tabel 6. Interpretasi Indeks Kesukaran Butir

Kriteria	Interpretasi
0,0 <i>logit</i> + 1SD	Soal sulit
> + 1SD	Soal sangat sulit
0,0 <i>logit</i> - 1SD	Soal mudah
< - 1SD	Soal sangat mudah

(Sumber: Sumintono dan Widhiarso, 2015)

### HASIL PENELITIAN DAN PEMBAHASAN

Pada tahapan desain soal, peneliti membuat 23 soal materi energi dan usaha, 17 soal materi cahaya dan 14 soal materi pesawat sederhana. Sukardi (2015) mengemukakan bahwa validitas adalah derajat yang menunjukkan kemampuan suatu tes mengukur apa yang hendak diukur. Validitas dilakukan secara teoritis dan empirik. Desain soal ini divalidasi oleh pakar dan praktisi. Validasi oleh pakar ini dilakukan dengan tujuan untuk menilai instrument soal yang dikembangkan sudah tepat secara teori dan sesuai dengan teori..Validasi desain dilakukan untuk mengetahui kelayakan instrumen kognitif yang dikembangkan kemudian dilakukan revisi berdasarkan saran-saran yang diberikan oleh validator. Validasi instrumen kognitif dilihat dari aspek konstruksi umum dan validasi butir. Instrumen penilaian kognitif yang telah dirancang oleh peneliti dilakukan validasi oleh pakar dan praktisi.

Tabel 7. Hasil Validasi Instrumen Penilaian

No	Materi	Hasil Validasi
1	Usaha dan Energi	3,4
2	Cahaya	3,67
3	Pesawat Sederhana	3,6

Instrumen yang sudah divalidasi diperbaiki berdasarkan saran dari validator agar diperoleh instrumen penilaian yang lebih baik. Menurut pakar desain instrument soal sudah valid namun perlu dilakukan perbaikan dari segi penulisan dan penggunaan kata agar mudah dipahami oleh pembaca. Setelah dilakukan perbaikan oleh peneliti, maka soal diujicobakan dalam skala kecil untuk menguji analisis butir soal. Skor yang diperoleh siswa pada uji coba produk dianalisis untuk mengetahui validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dari instrumen tersebut.

Tabel 8. Hasil validitas dan Reliabilitas

No	Kriteria	No Soal		
		Energi & Usaha	Pesawat Sederhana	Cahaya
1	Valid	3, 4, 5, 7, 8, 10, 12, 13, 14, 15, 16, 17, 18, 20, 2, 6, 11, 22	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14	1, 2, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17
2	Tidak Valid	9, 21, 23, 19	10	3, 4, 5, 8
Reliabilitas		0,96	0,91	0,89

Berdasarkan hasil ujicoba, soal yang memiliki validitas sangat rendah harus dibuang (tidak diujicoba lapangan) karena soal tersebut tidak memadai untuk mengukur apa yang hendak diukur. Dari hasil analisis terdapat 5 soal materi energi dan usaha tidak valid, 1 soal materi pesawat sederhana tidak valid dan pada materi cahaya terdapat 4 soal yang tidak valid. Soal yang tidak valid ini tidak dapat digunakan pada ujicoba luas, sehingga soal ini harus dibuang. Hal ini didukung dengan pernyataan Sumintono dan Widhiarso (2015) bahwa suatu tes haruslah valid artinya dapat mengukur sesuatu yang hendak diukur.

Tabel 9. Tingkat kesukaran butir soal

No	Kriteria	No Soal		
		Energi & Usaha	Pesawat Sederhana	Cahaya
1	Sukar	1, 2, 5, 6, 7, 9, 11, 16, 17, 19, 20, 21, 22, 23		10, 13, 14, 15
2	Sedang	3, 4, 8, 10, 12, 13, 14, 15, 18,	1, 2, 3, 4, 5, 7, 8, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 16, 17
3	Mudah		6, 9, 10, 12, 13	

Berdasarkan Tabel 9 dapat terlihat bahwa soal yang dibuat tidaklah terlalu mudah dan tidak terlalu susah. Soal tidak boleh semuanya sukar atau semuanya mudah. Soal yang dibuat terdiri atas soal yang sukar, sedang dan mudah. Kemudian soal dilakukan analisis daya pembeda. Hasilnya dapat dilihat pada Tabel 10.

Tabel 10. Perhitungan Daya Pembeda

No.	Kriteria	No. Soal		
		Energi & Usaha	Pesawat Sederhana	Cahaya
1	Dapat diterima	3, 4, 7, 8, 10, 12, 13, 14, 15, 17	-	16
2	Dapat diterima tapi perlu diperbaiki	5, 16	1, 3, 5, 6, 7, 9, 10, 11, 12, 13	1, 2, 6, 10, 12, 14, 17
3	Diperbaiki	18	2, 4, 8	

No.	Kriteria	No. Soal		
		Energi & Usaha	Pesawat Sederhana	Cahaya
4	Dibuang	1, 2, 6, 9, 11, 19, 20, 21, 22, 23	-	3, 4, 5, 7, 8, 9, 11, 14, 15

Berdasarkan hasil analisis butir soal juga didapatkan soal yang memiliki daya pembeda yang jelek. Soal yang memiliki daya pembeda yang jelek tidak dapat membedakan siswa kelompok atas dan kelompok bawah, sehingga soal tersebut tidak dapat digunakan.

Setelah dilakukan ujicoba kecil, dilakukanlah ujicoba luas dengan 200 siswa kelas VIII SMPN di Kota Banjarmasin. Hasil yang didapatkan dari hasil ujicoba lapangan dianalisis dengan menggunakan program Rasch. Data valid tidaknya soal diperoleh dari hasil pemrograman *Rasch*. Validitas pada pengembangan instrumen ini merujuk pada kesesuaian antara hasil alat ukur dengan kemampuan yang ingin diukur. Pembuktian adanya validitas pada pengembangan instrumen ini adalah dengan menggunakan nilai MNSQ, nilai ZSTD, dan nilai *Point Measure Correlation*. Adapun untuk butir-butir soal yang memiliki nilai MNSQ, ZSTD, dan *Pt Measure Corr* yang sesuai dengan kriteria sehingga tidak perlu diperbaiki. Menurut Sumintono dan Widhiarso (2015) butir soal dikatakan valid apabila nilai *outfit square* (MNSQ) memenuhi kriteria  $0,5 < \text{MNSQ} < 1,5$ , nilai *outfit Z-standar* (ZSTD) memenuhi kriteria  $- 2,0 < \text{ZSTD} < +2$ , dan nilai *Point Measure Correlation* (*Pt Mean Cor*) memenuhi  $0,4 < \text{Pt Mean Cor} < 0,85$ .

Tabel 11. Kesesuaian Item

No Soal	Outfit		Pt Measure	Kategori
	MNSQ	ZSTD	Corr	
Materi Energi dan Usaha				
3	Kurang bagus	Kurang Baik	Baik	Valid
8	Kurang bagus	Tidak dapat diprediksi	Baik	Valid
9	Kurang bagus	Data mempunyai perkiraan yang logis	Baik	Valid
1, 2, 5, 6,7, 12, 10,11,	Baik	Data mempunyai perkiraan yang logis	Baik	Valid
4	Baik	Terlalu mudah diprediksi	Baik	Valid
Materi Pesawat Sederhana				
12, 6, 8	Kurang bagus	Kurang Baik	Baik	Valid
2	Baik	Kurang Baik	Baik	Valid

No Soal	Outfit		Pt Measure Corr	Kategori
	MNSQ	ZSTD		
9, 10, 7, 11, 1, 4, 13	Baik	Data mempunyai perkiraan yang logis		
3 & 5	Baik	Kurang Baik Materi Cahaya	Baik	
8, 4, 1, 2, 3,	Baik	Data mempunyai perkiraan yang logis	Baik	Valid
7	Baik	Tidak dapat diprediksi	Baik	Valid
5, 6	Baik	Terlalu mudah diprediksi	Baik	Valid

Menurut Sumintono (2015) nilai kesesuaian butir yang terdiri dari *Outfit* MNSQ, ZSTD, dan *Pt Measure Corr* sangat dipengaruhi oleh besarnya ukuran sampel. Kesalahan menetapkan kunci jawaban, banyaknya individu yang kurang termotivasi dalam mengerjakan soal, atau soal yang memiliki daya beda rendah dapat menurunkan nilai ketepatan butir. Hasil penelitian Ardiyanti (2016), menyatakan bahwa hasil perhitungan validitas menggunakan model Rasch dapat dikatakan lebih akurat daripada perhitungan klasik karena data responden yang digunakan sudah bebas dari *outliner* atau responden yang tidak sesuai dengan pengukuran yang dilakukan. Namun, menurut Sugiyono (2015) dengan menggunakan instrumen yang sudah teruji validitasnya, bukan berarti data penelitian yang diperoleh pasti akan valid. Hal ini masih akan dipengaruhi oleh kondisi obyek yang diteliti, dan kemampuan orang yang menggunakan instrumen untuk mengumpulkan data. Oleh karena itu, peneliti harus mampu mengendalikan obyek yang diteliti dan meningkatkan kemampuan dalam menggunakan instrumen untuk mengukur variabel yang diteliti.

Reliabilitas yang berasal dari kata *reliability* berarti sejauh mana hasil suatu pengukuran dapat dipercaya. Hasil analisis program *Rasch* pada aspek reliabilitas untuk siswa dapat dilihat pada Tabel 8. Berdasarkan Tabel 8 tersebut, diperoleh rata-rata nilai seluruh siswa dalam mengerjakan butir-butir soal yang diberikan. Pada materi energi dan usaha didapatkan nilai *person reliability* adalah 0,77, yang dapat dikategorisasikan konsistensi jawaban siswa cukup. Sedangkan untuk nilai reliabilitas item sebesar 0,99 yang menyatakan bahwa kualitas butir-butir soal dalam instrument aspek reliabilitasnya istimewa. Adapun nilai alpha cronbach sebesar 0,72 artinya reliabilitas antara person dan butir soal secara keseluruhan bagus. Pada materi pesawat sederhana didapatkan bahwa nilai *person reliability*, yaitu sebesar 0,76 dan *item reliability*, yaitu sebesar 0,99 dengan kategori istimewa. Pada materi cahaya nilai rata-rata yang diperoleh sebesar -0.62 yang artinya lebih kecil dari 0.0 sehingga menunjukkan bahwa kemampuan yang dimiliki siswa lebih kecil daripada tingkat kesulitan soal. Reliabilitas siswa pada materi cahaya, yaitu sebesar 0.79 yang menunjukkan bahwa konsistensi jawaban dari siswa cukup bagus.

Artinya siswa memang dapat menjawab soal mudah dan kesulitan dalam menjawab soal yang sukar. Nilai reliabilitas (*Cronbach Alpha*) yang diperoleh sebesar 0.85 dan masuk dalam kategori bagus sekali, sehingga hasil pengukuran yang dilakukan dapat dipercaya. Nilai reliabilitas tersebut menunjukkan interaksi antara siswa dan butir-butir soal secara keseluruhan sangat bagus. Adapun nilai reliabilitas butir soal yang diperoleh adalah 0.99 yang masuk dalam kategori istimewa. Hal itu menunjukkan bahwa kualitas butir-butir soal yang digunakan dalam pengukuran sangat bagus.

Jadi, secara keseluruhan instrumen kognitif yang dikembangkan memiliki tingkat reliabilitas yang tinggi, baik dari segi reliabilitas individu, butir soal, maupun nilai *Alpha Cronbach*. Hasil penelitian Manfaat (2013) menyatakan bahwa jika nilai reliabilitas tes yang didapatkan tinggi maka tes yang dikembangkan dapat memberikan hasil yang sama bila diberikan pada kelompok yang sama meskipun oleh orang yang berbeda, waktu atau kesempatan yang berbeda, dan tempat yang berbeda pula sehingga konsistensi tes dianggap tinggi dan dapat dipercaya. Kereh (2015) menyatakan bahwa setidaknya ada dua sifat dasar yang dapat dijadikan acuan kualitas dari suatu instrumen pengukuran, yaitu reliabilitas dan validitasnya. Jika reliabilitas dan validitas instrumen kognitif yang dikembangkan itu valid, maka instrumen tersebut dapat digunakan seterusnya untuk melakukan pengukuran karena kualifikasinya yang baik.

Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks (Sudaryono, 2012). Tingkat kesulitan butir soal yang dikembangkan dapat diketahui melalui nilai logit yang ditunjukkan pada hasil program rasch. Pada materi energi dan usaha terdapat satu butir soal sangat sukar, lima soal sukar, tiga soal mudah dan tiga soal sangat mudah. Pada soal materi cahaya, berdasarkan hasil tersebut ada empat soal yang tergolong sulit, tiga butir soal mudah, dan satu butir soal sangat mudah. Pada materi pesawat sederhana terdapat tiga butir soal sangat sulit, satu butir soal sulit, dan enam butir soal dengan kategori mudah. Hal ini menunjukkan bahwa soal sudah variatif dari tingkat kesulitannya. Kereh (2015) menyatakan bahwa jika soal yang tergolong sulit bagi siswa memiliki jumlah lebih banyak daripada soal mudah, maka hal itu menegaskan kondisi siswa yang menjadi subyek penelitian ini termasuk dalam kelompok siswa dengan kemampuan penalaran yang kurang. Menurut Manfaat (2013), ada beberapa faktor yang menyebabkan butir soal berkategori sukar yaitu: (1) butir soal itu "mungkin" salah kunci jawaban; (2) butir soal itu mempunyai 2 atau lebih jawaban yang benar; (3) Materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya, sehingga kompetensi minimum yang harus dikuasai peserta didik belum tercapai; dan (4) pernyataan atau kalimat soal terlalu kompleks dan panjang. Namun, dari hasil penelitian soal yang dikembangkan memiliki kualitas yang baik karena memiliki tingkat kesukaran yang beragam, yaitu sangat sulit, sulit, mudah, dan sangat mudah.

## KESIMPULAN

Pada penelitian pengembangan ini dihasilkan instrumen kognitif berupa lima butir soal uraian pada materi cahaya, sembilan butir soal uraian materi pesawat

sederhana dan sembilan butir soal uraian materi energi dan usaha. Maka dapat diperoleh kesimpulan bahwa pengembangan instrumen penilaian kognitif fisika di kelas VIII SMP Kota Banjarmasin, dapat dinyatakan valid, reliabel, dan memiliki tingkat kesukaran yang bervariasi. Hal ini didukung oleh temuan sebagai berikut:

- 1) Validitas instrumen penilaian yang dikembangkan adalah valid berdasarkan hasil lembar validasi, uji coba produk, dan program Rasch (nilai *oufit* MNSQ, *oufit* ZSTD, dan korelasi pengukuran butir).
- 2) Reliabilitas instrumen penilaian yang dikembangkan adalah sangat reliabel berdasarkan hasil lembar validasi, uji coba produk, dan program Rasch.
- 3) Tingkat kesukaran dari instrument dengan program Rasch didapatkan tingkat kesukaran soal bervariasi, yaitu sangat sulit, sulit, mudah, dan sangat mudah.

#### DAFTAR PUSTAKA

- Ardiyanti, D. (2016) Aplikasi Model Rasch pada Pengembangan Skala Efikasi Diri dalam Pengambilan Keputusan Karier Siswa, *Jurnal Psikologi*, Vol 43, No 03
- Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Basuki, I. dan Hariyanto. (2015). *Asesmen Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Dirman dan Juarsih, C. (2014). *Penilaian dan Evaluasi*. Jakarta: PT Rineka Cipta.
- Kereh, C. T. (2015). Validitas dan Reliabilitas Instrumen Tes Matematika Dasar yang Berkaitan dengan Pendahuluan Fisika Inti. *Jurnal Pendidikan*. Vol 02, No 01
- Manfaat, B dan Siti N. (2013). Pengembangan Instrumen Tes untuk Mengukur Kemampuan Penalaran Statistik Mahasiswa Tadris Matematika. *Jurnal Jurusan Tadris Matematika*. Vol 02, No 02
- Rasyid, H. dan Mansyur. (2011). *Penilaian Hasil Belajar*. Bandung: CV Wacana Prima.
- Sudaryono. (2012). *Dasar-dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu
- Sugiyono. (2015). *Metode Penelitian dan pengembangan*. Bandung: Alfabeta.
- Sukardi, M. (2015). *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara.
- Sumintono, B. dan Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assesmen Pendidikan*. Cimahi: Trim Komunikata.
- Purwanto, M. N. (2013). *Prinsip-Prinsip dan teknik Evaluasi Pendidikan*. Bandung: PT Remaja Rosdakarya.
- Widyoko, E. P. (2013). *Evaluasi Program Pembelajaran*. Yogyakarta: Pustaka Belajar.