

ANALISIS KUALITAS BUTIR SOAL PADA MATA PELAJARAN FISIKA DI JURUSAN FISIKA FAKULTAS MIPA UNIVERSITAS NEGERI GORONTALO TAHUN AJARAN 2021/2022

***Nurfadillah Abdullah**
Universitas Negeri Gorontalo
nur871946@gmail.com

Mohammad Jahja
Universitas Negeri Gorontalo
mj@ung.ac.id

Dewa Gede Eka Setiawan
Universitas Negeri Gorontalo
dewaeka@ung.ac.id

Abstrak - Penelitian ini dilakukan untuk menganalisis kualitas butir soal yang digunakan untuk mengukur kemampuan mahasiswa pada mata pelajaran fisika di jurusan fisika, fakultas MIPA, Universitas Negeri Gorontalo. Peserta pada mata pelajaran ini terdiri dari 22 mahasiswa. Soal yang diberikan berupa soal pilihan ganda berjumlah 10 butir. Model rasch digunakan untuk mendapatkan butir soal yang fit. Analisis ini dilakukan dengan bantuan *software Winsteps*. Dari output program *Winsteps* diperoleh hasil 10 butir soal sesuai dengan model Rasch dengan nilai rata-rata Outfit MNSQ untuk person dan item masing-masing 1,26 dan 1,26. Sedangkan nilai Outfit ZSTD untuk person dan item masing-masing 0,07 dan 0,14. Sedangkan reliabilitas untuk person dan item masing-masing 0,50 dan 0,84.

*koresponden author

Kata Kunci : fit, model rasch, winsteps, reliabilitas

Abstract – This study was conducted to analyze the quality of the items used to measure students' abilities in physics subjects at the Department of Physics, Faculty of Mathematics and Natural Sciences, State University of Gorontalo. Participants in this subject consisted of 22 students. The questions given are in the form of multiple choice questions totaling 10 items. The Rasch model is used to obtain fit items. This analysis was carried out with the help of Winsteps software. From the output of the Winsteps program, 10 items were obtained according to the Rasch model with an average Outfit MNSQ value for persons and items of 1.26 and 1.26, respectively. While the Outfit ZSTD values for person and item are 0.07 and 0.14 respectively. While the reliability for the person and item respectively 0.50 and 0.84

Keywords: *fit, Rasch model, winsteps, reliability*

A. PENDAHULUAN

Menurut Undang-undang No. 20 tahun 2003, pendidikan merupakan usaha sadar dan terencana untuk mengembangkan segala potensi yang dimiliki peserta didik melalui proses pembelajaran. Mempunyai pengendalian diri, keterampilan, akhlak mulia kecerdasan kepribadian serta spiritual keagamaan yang berguna untuk kehidupan di lingkungannya maupun di masyarakat yang lebih luas merupakan tujuan dari diselenggarakannya pendidikan tersebut. Pembentukan karakter pada anak juga merupakan fungsi dari pendidikan. (UUD RIUU RI No. 28 Tahun 2004)

Kualitas pendidik di perguruan tinggi tentunya dipengaruhi oleh peran dosen dalam pembelajaran di kampus. Dengan berdasar pada tujuan dari pembelajaran yang ada, dosen kemudian mengarahkan proses pembelajaran sesuai dengan ketentuan tersebut. Tidak hanya sampai disitu, dosen pun patut untuk melakukan evaluasi terhadap peserta didiknya. Agar dosen tersebut dapat mengetahui hasil dari pembelajaran yang telah diberikan selama proses pembelajaran berlangsung. Kegiatan evaluasi merupakan sebuah proses aktivitas yang wajib ada dalam setiap penyelenggaraan aktivitas pembelajaran di setiap perguruan tinggi.

Evaluasi mempunyai masing-masing sudut pandangnya tersendiri tergantung dari berbagai keahlian dibidangnya. Orang-orang yang berada dalam bidang industry akan melakukan evaluasi terhadap hal-hal yang berkaitan dengan insdustri. Ahli yang menekuni pendidikan tentunya banyak bercerita tentang masalah pendidikan seperti kinerja, prestasi belajar, manajemen kelas, kompetensi guru, iklim akademis, dan seterusnya. Sementara itu seseorang yang menekuni bidang pendidikan akan banyak melakukan evaluasi terkait bidang pendidikan pula seperti prestasi belajar, iklim akademis, kinerja, dan lain sebagainya. (Mansyur, 2015)

Dalam menjalankan penilaian ruang lingkupnya ialah mengumpulkan bukti guna mendapatkan hasil capaian yang diperoleh oleh peserta didik. (Surapranata, 2004). Dalam kegiatan belajar mengajar penilaian capaian hasil belajar merupakan salah satu faktor utama didalamnya. Dalam upaya peningkatan pembelajaran, peningkatan kualitas sistem penilaian merupakan salah satu cara yang efisien. Semakin baik standarisasi dari penilaian yang ada akan menimbulkan dorongan bagi tenaga pengajar dalam memberikan pengajaran serta memunculkan motivasi untuk belajar yang lebih baik terhadap para peserta didik. (Widoyoko, 2011)

Adalah sebuah cara yang dipergunakan untuk mengukur seberapa tingginya tingkat kemampuan seseorang secara tidak langsung, yakni dengan melakukan metode mengajukan sejumlah pertanyaan dan jawaban dari responden dijadikan sebagai hasil dari tingkat kemampuan seseorang tersebut meupakan pendefinisian dari tes. Lebih jauh lagi dengan menggunakan standar numeric ataupun kategori, tes merupakan sebuah prosedur yang tepat untuk mendapatkan gambaran ataupun deskripsi mengenai karakteristik dari seseorang (Mardapi, 2008). Tes merupakan pertanyaan yang mempunyai kemungkinan jawaban benar atau salah. Berdasarkan fungsi, terdapat dua golongan dari tes, yakni tes awal atau yang disebut dengan pre-test serta tes terakhir atau yang disebut post-test. Pre test dijalankan

dengan cara pada saat sebelum materi diberikan guna menguji pemahaman mengenai materi yang akan di berikan kepada peserta didik. Serta untuk mengukur pengetahuan para peserta didik terkait bahan ajar yang akan diberikan. Sementara itu post-test dijalankan guna memperoleh pengetahuan mengenai tingkat pengetahuan yang dimiliki oleh peserta didik akan materi yang telah diberikan serta juga mendapatkan data mengenai peserta didik dalam kemampuannya yang diperoleh setelah proses pembelajaran selesai.(Mansyur, 2015)

Menjalankan analisa tes merupakan sebuah upaya agar mendapatkan data mengenai tingkat kualitas tes. Melakukan analisa soal memiliki tujuan agar soal-soal yang kurang baik, baik, serta tidak baik dapat teridentifikasi (Liesfi, 2016). Dalam melakukan analisa soal aspek yang harus terpenuhi ialah tingkat reliabilitas, praktibilitas, validitas, serta ekonomis (Arikunto, 2005).

Suatu butir soal, dikatakan telah menjalankan fungsi ukurnya ketika mampu untuk membedakan antara siswa yang mampu dan yang tidak mampu. Terdapat dua hal yang perlu dibahas lebih lanjut mengenai hal ini. Kemungkinan pertama adalah adanya kemungkinan ketidaksesuaian siswa/responden yang terlibat dalam ujian yang diberikan. Hal ini karena hasil yang diberikannya memang tidak sesuai dengan yang diperkirakan. Pemodelan Rasch dapat mendeteksi adanya siswa/responden yang memang tidak sesuai untuk dilibatkan berdasarkan respons yang dia berikan. Hal ini memberikan kelebihan pada pemodelan Rasch dengan ditemukannya alasan kenapa responden memang perlu dikeluarkan. Temuan ini jelas akan berkontribusi secara penting dalam analisis siswa dan soal yang dijalankan oleh guru sebagai penilai. Tidak seperti halnya dalam praktik pengolahan data statistik tradisional yang dilakukan tidak dapat diketahui secara pasti, namun dalam pemodelan Rasch hal ini akan muncul dengan sendirinya.(Sumintono & Widhiarso, 2015:45)

B. METODE

Penelitian ini dilakukan pada mahasiswa jurusan Fisika, Fakultas MIPA, Universitas Negeri Gorontalo tahun ajaran 2021/2022. Pengambilan data diambil pada saat Ujian Akhir Semester (UAS) semester Genap. Instrumen tes yang digunakan yaitu berupa butir soal tes objektif fisika. Data yang terkumpul dianalisis tingkat validitas dan reliabilitas dari instrumen butir tes yang digunakan. Hal ini dikarenakan dengan menggunakan model Rasch adalah solusi mengenai masalah validitas dimana model Rasch mampu menyediakan statistik serta menawarkan kesempatan untuk menyelidiki validitas instrumen tes dengan berdasarkan respon subjek penelitian.

Analisis validitas yang dihasilkan dengan program *Winstep*® yaitu analisis berupa validitas konstruk dan validitas konten, analisis validitas konten meliputi tingkat kesesuaian soal yang berfungsi untuk melihat kualitas tingkat kesesuaian butir dengan model. Informasi yang diberikan berupa informasi kesesuaian butir soal dengan kriteria yaitu dengan melihat nilai outfit *mean square*, outfit *z-standard*, dan point *measure correlation*. Butir soal dikatakan valid atau diterima jika telah memenuhi minimal 2 kriteria dan diperbaiki jika memenuhi salah satu dari ketiga kriteria tersebut, serta dibuang

bila tidak ada yang memenuhi kriteria tersebut. Nilai kesesuaian butir sangat dipengaruhi oleh banyaknya data, semakin besar sampel yang digunakan maka semakin baik juga tingkat kesesuaiannya.

Nilai kriteria yang dipakai dalam melihat tingkat kesesuaian butir (Sumintono & Widhiarso, 2015):

- a. Nilai Outfit *Mean Square* (MNSQ) yang diterima: $0,5 < \text{MNSQ} < 1,5$
- b. Nilai Outfit *Z-Standard* (ZSTD) yang diterima: $-2,0 < \text{ZSTD} < +2,0$
- c. Nilai *Point Measure Correlation* (Pt Measure Corr): $0,4 < \text{Pt Measure Corr} < 0,85$

Sumintono & Widhiarso, (2015) menyatakan bahwa pemodelan Rasch dapat melakukan analisis sampai ke tingkat butir soal, tingkat individu, dan tingkat instrument secara lebih rinci dalam hal ini akan membimbing guru dalam pembuatan soal, ataupun dalam mengambil kebijakan yang tepat, logis, dan ilmiah berdasarkan panduan kriteria kualitas instrumen sebagai berikut:

- a. Person dan item *measure* menunjukkan nilai rata-rata peserta didik dan item, nilai yang paling kecil menunjukkan tingkat kemampuan peserta didik dan soal yang rendah dengan nilai logit 0,0.
- b. Nilai *alpha crombach* nilai ini digunakan untuk mengukur reliabilitas yakni interaksi antara soal dan item, kriteria yang digunakan yakni ($< 0,5$: Buruk, $0,5 - 0,6$: Jelek, $0,6 - 0,7$: Cukup, $0,7-0,8$: Bagus, $> 0,8$: Istimewa).
- c. Nilai person *reliability* dan item *reliability* merupakan nilai reliabilitas peserta didik dan nilai reliabilitas item soal. Kategori yang digunakan yaitu: $< 0,67$: Lemah, $0,67 - 0,80$: Cukup, $0,8 - 0,90$: Bagus, $0,91 - 0,94$: Bagus sekali, $> 0,94$: Istimewa.
- d. *Separation* merupakan pengelompokan peserta didik dan item. Semakin besar nilai *separation* maka kualitas instrumen akan semakin baik. Persamaan (1) menunjukkan persamaan yang digunakan dalam melihat pengelompokkan peserta didik dan item.

$$H = \frac{[(4 \times \text{Separation})+1]}{3} \quad (1)$$

Keterangan:

H : Nilai strata (Pengelompokan)

Separation : Nilai separation untuk responden yang dihasilkan

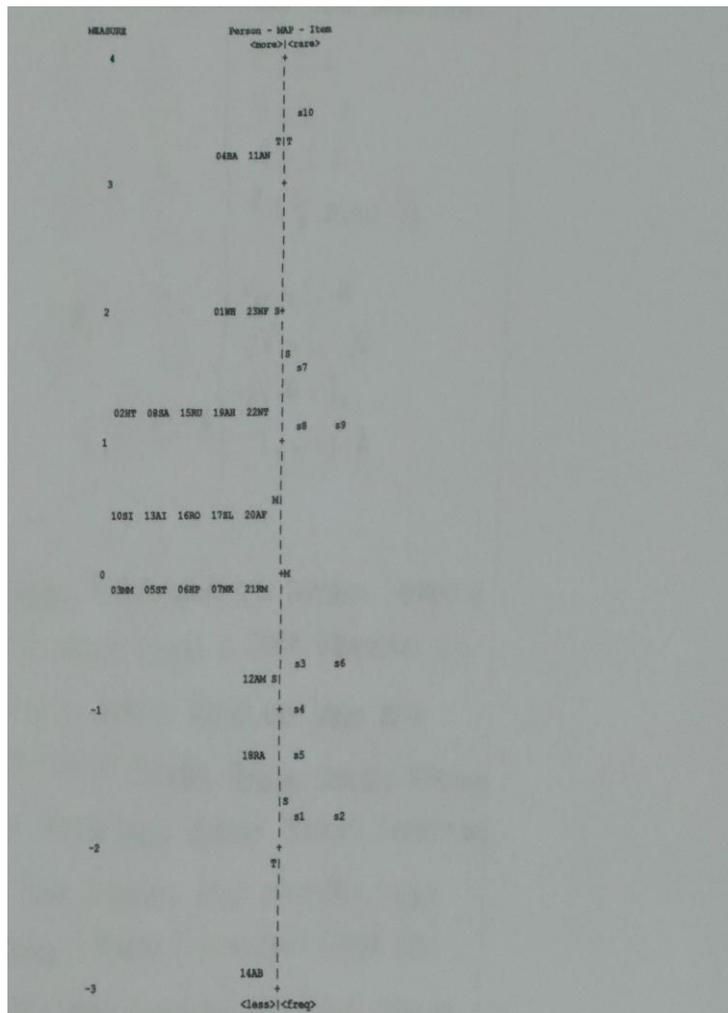
Tingkat kesukaran butir soal (item *measure*) memberikan tampilan nilai logit dari setiap butir soal secara terperinci. Hasil output dari tabel ini memberikan informasi butir soal yang diurutkan dari yang memiliki nilai logit *measure* yang tertinggi ke nilai logit yang terendah yang menunjukkan bahwa soal diurutkan dari butir soal yang tersukar sampai pada yang butir soal yang termudah. Pengelompokkan tingkat kesukaran soal dapat ditentukan berdasarkan nilai rata-rata logit yang dijumlahkan dengan nilai standar deviasi (SD). Nilai ini berguna untuk mengidentifikasi kelompok item (*separation*).

C. HASIL DAN PEMBAHASAN

Perangkat lunak (*software*) *Ministep* adalah program komputer khusus analisis pemodelan yang bisa bekerja di bawah sistem Microsoft Windows yang dibuat oleh John Linacre. Pengolahan data dengan menggunakan *Ministep* memerlukan jenis berkas/*file* tertentu yang isinya hanya berbentuk data mentah saja. File data yang digunakan bisa melalui berbagai cara, namun untuk mudahnya digunakan program lain yang sudah biasa dikenal dan praktis digunakan. Karena itu penyiapan berkas data mentah digunakan perangkat lunak *spreadsheet* Microsoft Excel (Sumintono & Widhiarso, 2015).

Salah satu keistimewaan pemodelan Rasch dengan program *Ministep* adalah menghasilkan suatu peta yang menggambarkan sebaran kemampuan siswa/responden dan sebaran tingkat kesulitan soal dengan *skala yang sama*. Peta ini biasa disebut Wright Map (Peta Wright, diambil dari pencetusnya Benjamin Wright) yang tidak lain adalah peta *person-item* secara komprehensif. Gambaran sebaran kemampuan 22 (dua puluh dua) mahasiswa dan sebaran kesukaran butir soal pada skala yang sama. Hasil analisis data pada gambar 1. Menunjukkan bahwa ada dua mahasiswa memiliki kemampuan tinggi yaitu mahasiswa dengan kode 04BA & 11AN dengan nilai *logit* yang dicapai +3, hal ini menunjukkan bahwa ada beberapa mahasiswa sudah memahami mengenai konsep materi yang diujikan, dari data juga dapat diketahui bahwa rata-rata mahasiswa berada pada kemampuan menengah dengan nilai *logit* +2 dan +1, dan ada juga satu orang mahasiswa dengan kemampuan rendah dengan nilai *logit* -3 yang artinya masih ada yang belum memahami konsep materi yang diujikan.

Pada gambar 1, juga dapat kita lihat bahwa ada satu soal yang memiliki nilai kesukaran yang tinggi yaitu pada soal nomor s10 Adapun soal yang memiliki nilai kesukaran menengah seperti pada soal s3 dengan nilai *logit* -0 dan +0, dan beberapa soal yang memiliki tingkat kesukaran yang rendah dengan nilai *logit* dibawah -1, soal-soal yang memiliki nilai *logit* -1 dibawah inilah yang harus direvisi kembali. Hasil ini sesuai menurut Boopathiraj dan Chellamani (2013) mengemukakan bahwa soal yang memiliki daya beda tinggi adalah soal dimana peserta didik yang memiliki skor tes rendah tidak dapat menjawab soal dengan benar.



Gambar 1. Peta Wright

Penjelasan sebelumnya membuktikan bahwa masih ada beberapa mahasiswa yang belum memahami konsep Fisika yang diujikan, sebagai calon guru fisika sebaiknya hal seperti ini dapat bias diminimalisir dengan memahami konsep-konsep fisika yang sebenarnya menurut para ahli agar siswa-siswa di masa depan dapat memahami materi-materi fisika.

1. Validitas Butir Soal

Berdasarkan hasil analisis nilai *Outfit Mean Squared* (*Outfit MNSQ*) sebesar 1,26 baik pada kolom person maupun item. Nilai 1,26 termasuk dalam kriteria fit yaitu terletak diantara selang $0,5 < MNSQ < 1,5$, yang artinya instrumen tes yang digunakan sudah sesuai dengan model untuk mengukur kompetensi mahasiswa pada materi yang diujikan. Selanjutnya, diperoleh nilai *Outfit Z Standardized* (*Outfit ZSTD*) sebesar 0,07 untuk person 0,14 untuk item. Nilai 0,07 dan 0,14 berada diantara selang $-2,0 < MNSQ < 2,0$ yang dapat diartikan data memiliki kemungkinan nilai rasional. Hal ini berarti secara keseluruhan butir soal atau item telah sesuai dengan model rasch dan dapat dijadikan instrumen tes prestasi pada materi yang diujikan.

Distribusi item soal yang dianggap *misfit* atau tidak *fit* dengan model dapat dilihat pada Tabel 2. Batas item dinyatakan *fit* dengan model jika memenuhi salah satu atau keduanya dari syarat berikut. Syarat pertama, nilai Outfit MNSQ terletak diantara 0,5 sampai dengan 1,5; nilai Outfit ZSTD terletak diantara -2,0 sampai dengan 2,0; dan nilai korerasi butir soal dengan skor total (point measure correlation) terletak diantara 0,4 sampai dengan 0,85. (Sumintono & Widhiarso, 2015)

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		Item
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
10	2	22	3.52	.80	1.56	1.06	5.21	2.25	A-.24	.33	90.9	90.8	s10
8	9	22	1.07	.50	1.21	1.01	1.43	1.00	B .31	.47	63.6	71.5	s8
1	19	22	-1.78	.70	1.01	.19	1.24	.56	C .36	.41	90.9	88.7	s1
2	19	22	-1.78	.70	1.17	.49	.65	-.04	D .39	.41	81.8	88.7	s2
9	9	22	1.07	.50	1.15	.76	1.11	.40	E .38	.47	63.6	71.5	s9
4	17	22	-.98	.58	.85	-.39	.57	-.54	e .57	.44	77.3	81.4	s4
7	7	22	1.58	.52	.85	-.54	.76	-.29	d .56	.46	77.3	76.2	s7
5	18	22	-1.35	.63	.81	-.40	.62	-.27	c .55	.43	90.9	84.8	s5
6	16	22	-.67	.54	.73	-1.03	.55	-.75	b .63	.45	86.4	78.0	s6
3	16	22	-.67	.54	.65	-1.40	.47	-.97	a .68	.45	86.4	78.0	s3
MEAN	13.2	22.0	.00	.60	1.00	-.02	1.26	.14			80.9	81.0	
P.SD	5.7	.0	1.65	.10	.26	.81	1.35	.91			9.9	6.7	

Gambar 2. Tingkat Kesesuaian Butir Soal

Berdasarkan hasil analisis instrumen tes prestasi dengan menggunakan bantuan program winsteps version 4,4.5 yang ada pada gambar 2 diperoleh butir soal sebanyak 10 butir, dan soal fit sebanyak 10 butir, sehingga diperoleh instrumen final sebanyak 10 butir.

2. Reliabilitas Instrumen

Tabel 2 Reliabilitas Instrumen

Variabel	Rata-rata logit (SD)	Separation	Reliability
Mahasiswa	0,89	1,00	0,50
Soal	1,36	2,32	0,84

Dari hasil analisis diperoleh bahwa rata-rata nilai seluruh mahasiswa dalam mengerjakan soal yang diberikan adalah 0,89 *logit*. Nilai tersebut lebih besar dari nilai rata-rata kesulitan artinya terdapat kecenderungan mahasiswa yang lebih tinggi dari pada tingkat kesulitan soal. Mahasiswa mendapatkan nilai 1,00 dan soal diperoleh 2,32. Dapat dikatakan bahwa semakin besar nilai separation, maka mampu menunjukkan kualitas instrumen butir soal yang digunakan sangat bagus. Hal ini dikarenakan mampu mengidentifikasi kelompok butir soal dengan kelompok responden.

$$H = \frac{[(4 \times \text{Separation}) + 1]}{3}$$
$$H_{\text{mahasiswa}} = \frac{[(4 \times 1,00) + 1]}{3} = 1,66$$
$$H_{\text{soal}} = \frac{[(4 \times 2,32) + 1]}{3} = 3,42$$

Mulai perhitungan diperoleh $H_{\text{mahasiswa}}$ ada 1 sedangkan H_{soal} ada 3 maka pengelompokkan yang bias dilakukan maksimal 3 karena indikasi separasi yang ada. Reliabilitas mahasiswa diperoleh hasil 0,50 berada pada kategori lemah dan reliabilitas soal diperoleh hasil 2,32 bagus.

D. SIMPULAN

Berdasarkan hasil dari analisis data dan pembahasan diperoleh kesimpulan hasil analisis menggunakan program *Winstep* memberikan informasi, baik dari segi item maupun responden menunjukkan perbedaan terhadap butir soal dan mahasiswa yang dianalisis menggunakan model rasch.

1. Hal ini ditunjukkan dengan skor item (*item reliability*) sebesar 0,84, dan separation 2,32. Sedangkan reliabilitas person (*person reliability*) sebesar 0,50, dan separation 1,00.
2. Dan dapat kita lihat juga bahwa ada satu soal yang memiliki nilai kesukaran yang tinggi yaitu pada soal nomor 10.
3. Adapun soal yang memiliki nilai kesukaran menengah seperti pada soal s3 dengan nilai logit -0 dan +0, dan beberapa soal yang memiliki tingkat kesukaran yang rendah dengan nilai logit dibawah -1, soal-soal yang memiliki nilai logit -1 dibawah inilah yang harus direvisi kembali.

DAFTAR RUJUKAN

- Arikunto, Suharsimi. 2006. *Dasara-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- . 2009. *Dasar-dasar Pendidikan*. Jakarta: PT. Bumi Aksara.
- Azizah & Sapti Wahyuningsih. 2020. "Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria" :1-50.
- Bambang, Subali. (2012). *Prinsip Asesmen & Evaluasi Pembelajaran*. Yogyakarta: UNY Press.
- Daryanto. 2012. *Evaluasi Pendidikan*. Jakarta: Renika Cipta.
- Liesfi, Zaenal Arifin dan. 2010. *Evaluasi Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Mansyur, Rasyid, H., & Suratno. (2009). *Assesmen pembelajaran di sekolah*. Yogyakarta: Multi Pressindo.
- Mansyur, Rasyid, H., & Suratno. (2015). *Assesmen pembelajaran disekolah. Panduan bagi guru dan calon guru*. Yogyakarta: Pustaka Pelajar.
- Surapranata, S. (2004). *Panduan Penulisan Tes Tertulis*. Bandung: Rosda Karya.
- Suroso. 2016. *Analisis Kesalahan Siswa dalam Mengerjakan Soal-soal Fisika Termodinamika Pada Siswa SMS Negeri 1 Magetan*. Jurnal Edukasi Matematika dan Sains (JEMS), 4 (1), 8-18.

Sumintono, Bambang. *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: implementasi Penilaian Formatif*.

UUD RI No. 20 Tahun. 2003. *Tentang Sistem Pendidikan Nasional*

UUD RIUU RI No. 28 Tahun. 2004. "Presiden republik indonesia." *Peraturan Pemerintah Republik Indonesia Nomor 26 Tahun 1985 Tentang Jalan 2003(1): 1–5.*

Widoyoko, S. E. P. (2011). *Evaluasi program pembelajaran*. Yogyakarta: Pustaka Pelajar