



Analisis butir dan deteksi bias soal fisika pada ujian sekolah berstandar nasional (USBN) dengan menggunakan pemodelan *rasch* untuk standarisasi penilaian

Nita Andriani*, Endi Suhendi, Achmad Samsudin

Magister Pendidikan Fisika, Universitas Pendidikan Indonesia Bandung, Indonesia

*e-mail: nitaandriani@upi.edu

Abstrak

Penelitian ini bertujuan untuk memperoleh hasil kajian tentang kualitas butir soal mata pelajaran Fisika pada Ujian Sekolah Berstandar Nasional (USBN) siswa SMK di Siak tahun 2018. Data yang digunakan adalah data hasil USBN Fisika siswa salah satu SMK di Kabupaten Siak Propinsi Riau pada USBN tahun 2018 sebanyak 20 orang siswa (studi kasus). Dalam analisis ini, 40 soal USBN dikelompokkan ke dalam dua kelompok yaitu 24 soal kategori *LOTS* dan 16 soal kategori *HOTS*. Hasil analisis dengan *rasch* model menunjukkan bahwa *person reliability* pada soal USBN kelompok *LOTS* adalah sebesar 0,04 dan pada soal kelompok *HOTS* sebesar 0,29. Sedangkan *item reliability* pada soal kelompok *LOTS* adalah 0,83 dan kelompok *HOTS* yaitu 0,74. Dari 40 soal yang dianalisis terdeteksi satu soal yang terjangkiti *DIF* atau bias. Penggunaan analisis dengan pemodelan *Rasch* dianjurkan untuk digunakan oleh setiap guru agar dapat membantu guru dalam memetakan kemampuan peserta didik dari analisis soal yang diberikan.

Kata kunci : analisis butir soal, model *rasch*, penilaian usbn fisika

1. Pendahuluan

Salah satu bagian penting dalam proses pembelajaran adalah penilaian. Penilaian merupakan penerapan berbagai cara dan penggunaan beragam alat penilaian dalam rangka memperoleh informasi untuk membuat keputusan tentang peserta didik, kurikulum, sekolah dan kebijakan pendidikan (Nitko & Brookhart, 2007). Dengan melakukan penilaian, seorang pendidik (guru) dapat mengetahui kemampuan yang dimiliki peserta didiknya serta ketepatan metode mengajar yang digunakan untuk kebaikan proses belajar selanjutnya (Marjiasih & Wahyuni, 2014). Hal ini tertuang dalam Permendikbud nomor 23 tahun 2016 pasal 4 ayat (1) tentang tujuan penilaian : *"bahwa penilaian hasil belajar oleh pendidik bertujuan untuk memantau dan mengevaluasi proses, kemajuan belajar, dan perbaikan hasil belajar peserta didik secara berkesinambungan"*.

Penilaian hasil belajar akan memberikan gambaran informasi dari tiap-tiap peserta didik tentang prestasi dan pencapaian kompetensi yang diperoleh dari proses pembelajaran Rosana & Sukardiyono, 2015). Penilaian meliputi tiga

aspek penting yaitu aspek pengetahuan, sikap dan keterampilan (Safitri, Sari & Wahyuni, 2017). Upaya untuk meningkatkan kualitas pendidikan dapat ditempuh melalui peningkatan proses pembelajaran dan kualitas sistem penilaiannya. Salah satu penilaian yang dilakukan di sekolah adalah penilaian sumatif. Yaitu penilaian yang dilakukan untuk mengetahui apa yang sudah diketahui oleh peserta didik atau yang bisa dia lakukan pada periode akhir masa belajar yang telah ditetapkan (Chiang, 2015).

Kusuma (2019) menyebutkan bahwa hasil belajar digambarkan sebagai pernyataan tertulis tentang apa yang diharapkan diketahui, dipahami dan atau dapat dilakukan peserta didik pada akhir periode pembelajaran. Tujuannya untuk memberikan informasi tentang prestasi yang telah dicapai, dalam istilah populernya disebut *assessment of learning*. Namun penilaian hasil belajar ini masih berfokus pada pemenuhan dimensi kognitif saja, masih sangat sedikit yang melatih pada keterampilan peserta didik (Putra, Masril, & Yurnetti, 2018). Pada jenis penilaian ini peserta didik berada dalam suatu situasi dimana mereka harus menampilkan segala

yang telah dikuasai selama waktu tertentu yang menunjukkan prestasi belajarnya, seperti Ujian Nasional (Putri & Istiyono, 2017). Untuk sekolah menengah kejuruan, mata pelajaran fisika tidak termasuk kedalam kelompok mata pelajaran yang di Ujian Nasionalkan, melainkan termasuk kedalam kelompok mata pelajaran USBN (Ujian Sekolah Berstandar Nasional).

Ujian sekolah berstandar nasional ini biasanya disusun oleh kumpulan guru sebidang di tiap kabupaten atau propinsi. Dengan mengacu kepada standar kompetensi lulusan tiap jurusan di masing-masing sekolah.

Hasil yang diperoleh siswa dari ujian sekolah ini menjadi gambaran bagaimana penguasaan mereka terhadap materi pelajaran. Namun adakalanya skor yang diperoleh peserta didik tidak sesuai dengan kemampuan yang sebenarnya. Penyebabnya bisa diakibatkan oleh permasalahan yang muncul dari peserta didik itu sendiri atau bisa juga disebabkan oleh kualitas butir soal yang diberikan (Rosana & Sukardiyono, 2015).

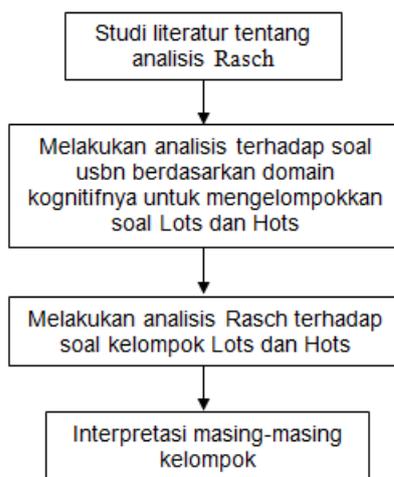
Untuk penilaian hasil belajar, sudah banyak model analisis yang dilakukan. Selain teori tes klasik dan teori respon butir, ada model analisis hasil penilaian yang populer yaitu model Rasch. Model Rasch menggunakan satu parameter logistic dalam menganalisis kemampuan responden (peserta didik), dengan menggunakan aplikasi *winstep*. Analisis dengan model Rasch cukup mudah dilakukan namun menghasilkan hasil analisis yang akurat. Rasch meninjau peluang menjawab benar pada soal bentuk dikotomi dengan membandingkan kemampuan peserta didik dengan tingkat kesukaran soal (Sumintono dan Widhiarso, 2014). Rasch model memiliki beberapa kelebihan yaitu mengakomodasi pendekatan probabilitas

dalam memandang atribut seluruh objek yang diukur, mengatasi masalah perbedaan metric antar butir, cukup tahan terhadap data hilang, dan telah memenuhi pengukuran yang objektif. Berdasarkan penjelasan inilah maka analisis dalam penelitian ini menggunakan pemodelan Rasch. Analisis Rasch juga memberikan pengembangan dan efisiensi terhadap instrument dengan tipe multiple choice (Hofer, 2017).

2. Metode

Populasi pada penelitian ini adalah seluruh lembar jawaban usbn siswa kelas XII salah satu SMK di kabupaten Siak pada tahun 2018. Sedangkan sampelnya adalah ruang 3 yang berjumlah 20 lembar jawaban usbn fisika siswa. Dari 20 lembar ini terdapat 11 lembar jawaban yang merupakan siswa kelas komputer dan jaringan, sedangkan 9 lembar lainnya berasal dari siswa kelas pemesinan. Kajian dalam penelitian ini menggunakan pemodelan Rasch untuk menganalisis item dan responden. Model Rasch dipilih karena berbagai kelebihan yang telah disampaikan pada pendahuluan. Solekhah (2018) menyebutkan bahwa Rasch model dapat memberikan analisis yang lebih lengkap untuk instrumen tes berpikir tingkat tinggi. Pengembangan instrumen penilaian yang relevan dan sesuai standar menjadi salah satu tugas pendidik (Widowati, 2016). Untuk mendukung analisisnya, digunakan sebuah *software* yaitu *ministep*. Instrumen soal yang dianalisis berjumlah 40 soal. Dari analisis data ini diharapkan dapat memberikan gambaran tentang analisis kemampuan siswa SMK tersebut terhadap soal usbn fisika berdasarkan model Rasch.

Desain penelitian yang dilakukan yaitu seperti gambar berikut :



3. Hasil dan Pembahasan

Data yang diperoleh dalam kajian ini merupakan data hasil usbn fisika siswa yang terdiri dari 40 soal pilihan ganda. Soal dikelompokkan dalam kelompok *LOTS* (*Lower Other Thinking Skills*) dan kelompok *HOTS* (*Higher Other Thinking Skills*).

a. Analisis butir

Menurut Ratumanan & Laurens dalam Palimbong (2018), disebutkan bahwa proses analisis butir soal merupakan kegiatan yang dilakukan untuk menelaah

dan mengkaji setiap butir soal melalui pengumpulan informasi dari jawaban peserta didik agar diperoleh soal yang bermutu sebelum soal-soal tersebut digunakan. Analisis butir meliputi :

1. Nilai Reliabilitas

Reliabilitas suatu instrument menunjukkan seberapa jauh pengukuran yang dilakukan berkali-kali akan menghasilkan informasi yang sama (Sumintono & Widhiarso, 2014).

Tabel 1. Perbandingan nilai reliabilitas kedua kelompok

Analisis	Kelompok Lots	Kelompok Hots
<i>Person reliability</i>	0.04	0.29
<i>Item Reliability</i>	0.83	0.74

Tabel di atas menunjukkan bahwa untuk kedua kelompok soal nilai *person reliability*-nya lemah, sedangkan *item reliability* termasuk kategori bagus untuk kelompok *Lots* dan cukup untuk kelompok *Hots*. Nilai *alpha cronbach* untuk kedua kelompok soal adalah bagus sekali yaitu 1.68 untuk kelompok *Lots* dan 1.56 untuk soal kelompok *Hots*. *Person Measure* menunjukkan nilai rata-rata responden dalam instrumen. Nilai rata-rata yang lebih dari logit 0.0 menunjukkan kecenderungan responden yang lebih banyak menjawab benar pada berbagai item soal.

2. Tingkat kesukaran butir soal

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

soal. Distribusi tingkat kesukaran dan kemampuan responden (peserta didik) dapat dilihat pada *Person-Item Map*. Bentangan tingkat kesukaran soal dan tingkat kemampuan peserta didik tersebut berada pada satu garis sehingga dapat diketahui posisi setiap peserta didik terhadap tingkat kesulitan butir yang dikerjakan. Tingkat kemampuan responden maupun tingkat kesukaran butir dalam *Rasch Model* diekspresikan dalam satu garis berupa absis pada grafik dengan satuan berupa *logit* (*logg-odd unit*).

Tingkat kesulitan butir pada model Rasch pada dasarnya sama dengan taraf kesukaran teori tes klasik, yaitu perbandingan antara jumlah jawaban benar dengan jumlah soal yang diujikan (*odd-ratio*). Hanya saja yang

membedakan adalah, nilai peluang itu kemudian diskalakan dengan memasukkan fungsi logaritma. Hasil estimasi logit dari *odd-ratio* inilah yang disebut logit atau *W-score* atau nilai *measure*. Jika pada teori tes klasik nilai indeks kesukaran yang tinggi berarti soal tersebut mudah, pada Rasch model nilai logit yang tinggi menunjukkan item tersebut sulit. Sama seperti dalam teori tes klasik, tidak ada patokan berapa tingkat kesulitan yang diterima dalam tes. Hal ini bergantung dari tujuan tes itu sendiri.

Sebaran tingkat kesulitan butir ini disajikan dalam bentuk peta.

Peta ini menyajikan sebaran responden menurut tingkat kemampuannya dan sebaran butir menurut tingkat kesukarannya. Untuk kelompok soal Lots yang berjumlah 24 soal, terdapat 2 soal berkategori mudah, 18 soal kategori sedang dan 4 soal kategori sukar. Sedangkan untuk kelompok soal Hots yang berjumlah 16 soal, didapati 1 soal kategori mudah, 14 soal kategori sedang dan 1 soal kategori sukar.

Tabel 2. Perbandingan tingkat kesukaran soal

Kelompok Soal	Kategori		
	Mudah	Sedang	Sukar
Lots	2	18	4
Hots	1	14	1

Dari tabel terlihat soal dengan kategori sukar pada kelompok soal Lots lebih banyak daripada soal sukar pada kelompok Hots. Ini menunjukkan bahwa soal dengan domain kognitif C1-C3 bisa jadi sangat sulit bagi siswa dibandingkan dengan soal pada domain kognitif C4-C6.

Butir soal yang paling sukar dalam kelompok soal Lots yaitu soal nomor 2 yang berada pada bagian teratas peta dengan tingkat kesukaran 3,49 dan yang paling mudah yaitu soal nomor 16 dengan tingkat kesukaran -3,25. Sedangkan untuk kelompok soal Hots, soal yang paling sukar yaitu soal nomor 34 dengan tingkat kesukaran 3,25 dan soal yang paling mudah yaitu soal nomor 25 dengan tingkat kesukaran -1,96.

Untuk memeriksa item yang tidak sesuai (*outliers* atau *misfits*) adalah dengan memeriksa hal-hal berikut (Sumintono & Widhiarso; 2014): Nilai INFIT MNSQ dari setiap item; yaitu dengan membandingkan nilai infit MNSQ setiap item dengan jumlah nilai rata-rata dan standar deviasi. Nilai logit yang lebih besar dari penjumlahan rata-rata dan standar deviasi mengindikasikan item yang misfit. Kriteria lain yang biasa juga digunakan untuk memeriksa item yang misfit yaitu:

- Nilai *Outfit Mean Square (MNSQ)* yang diterima : $0.5 < MNSQ < 1.5$
- Nilai *Outfit Z-Standard (ZSTD)* yang diterima : $-2.0 < ZSTD < +2.0$
- Nilai *Point Measure Correlation (Pt Mean Corr)* : $0.4 < Pt Mean Corr < 0.85$

Dari hasil analisis diatas, untuk kelompok soal Lots semua soal adalah *fit* begitu juga dengan kelompok soal Hots semua item *fit*, artinya tidak ada soal yang perlu direvisi atau di eliminasi. Dari analisis kualitas butir soal pilihan ganda USBN mata pelajaran Fisika SMK tahun 2018 ini secara kuantitatif baik menurut pemodelan *Rasch*.

3. Unidimensionalitas
Unidimensionalitas instrumen adalah ukuran yang penting untuk mengevaluasi apakah instrumen yang dikembangkan mampu mengukur apa yang seharusnya diukur. *Rasch Model* menggunakan analisis komponen utama (*Principal Component Analysis*) dari residual, yaitu mengukur sejauh mana keragaman dari instrumen mengukur apa yang seharusnya diukur.

Unidimensionalitas instrumen terpenuhi jika nilai *raw varians* nya minimal 20% dengan kategori cukup, apabila nilainya besar dari 40% artinya bagus dan jika besar dari 60% artinya istimewa (Sumintono & Widhiarso, 2014).

Tabel 3. Raw variance untuk soal Lots

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed	Expected
Total raw variance in observations	= 41.3737	100.0%	100.0%
Raw variance explained by measures	= 18.3738	44.4%	44.5%
Raw variance explained by persons	= 2.7254	6.6%	6.6%
Raw Variance explained by items	= 15.6484	37.8%	37.9%
Raw unexplained variance (total)	= 23.0000	55.6%	55.5%
Unexplned variance in 1st contrast	= 4.4974	10.9%	19.6%
Unexplned variance in 2nd contrast	= 3.4497	8.3%	15.0%
Unexplned variance in 3rd contrast	= 2.7959	6.8%	12.2%
Unexplned variance in 4th contrast	= 2.2675	5.5%	9.9%
Unexplned variance in 5th contrast	= 1.6721	4.0%	7.3%

Tabel 4. Raw variance untuk soal Hots

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed	Expected
Total raw variance in observations	= 23.2505	100.0%	100.0%
Raw variance explained by measures	= 7.2505	31.2%	31.1%
Raw variance explained by persons	= 1.9523	8.4%	8.4%
Raw Variance explained by items	= 5.2982	22.8%	22.7%
Raw unexplained variance (total)	= 16.0000	68.8%	68.9%
Unexplned variance in 1st contrast	= 3.6176	15.6%	22.6%
Unexplned variance in 2nd contrast	= 2.6091	11.2%	16.3%
Unexplned variance in 3rd contrast	= 2.0591	8.9%	12.9%
Unexplned variance in 4th contrast	= 1.7186	7.4%	10.7%
Unexplned variance in 5th contrast	= 1.2434	5.3%	7.8%

Dari soal *USBN* yang dianalisis, untuk kelompok soal Lots hasil pengukuran *raw varians* nya yaitu 44,4 % artinya bagus, sedangkan untuk kelompok soal Hots *raw varians* nya 31,2 % artinya cukup. Dengan demikian secara umum instrumen tes soal Ujian Sekolah Berstandar nasional (*USBN*) SMK ini cukup valid untuk mengukur kemampuan peserta didik.

b. Deteksi Bias

Deteksi bias pada item dalam analisis *Rasch Model* ditampilkan dalam keberfungsian item diferensial (*Differential Item Functioning* atau *DIF*). Hal ini diperlukan untuk mengetahui apakah item-item yang diberikan mempunyai bias dalam kategori responden tertentu atau tidak (Sumintono & Widhiarso, 2014). Dalam instrumen yang digunakan, dimasukkan dua data demografi yaitu jenis kelamin (laki-laki (L) dan perempuan (P)) serta kelas/jurusan (komputer-jaringan (K) dan mesin (M)). Bias dalam item dapat diketahui berdasarkan nilai probabilitas item yang berada di bawah 5 % (0,05).

Untuk instrumen soal *USBN* Fisika pada kelompok soal Lots, terdapat satu soal

yang bias terhadap jender yaitu soal nomor 30 dengan nilai probabilitas 0,0279. Sedangkan untuk kelompok soal Hots tidak ada soal yang terjangkiti *DIF* atau bias, baik terhadap jender maupun kelas.

Bias pada butir sebenarnya bukanlah karakteristik yang dijadikan pertimbangan utama dalam seleksi *item*. Meskipun demikian informasi mengenai adanya *item* yang bias sangat berpengaruh terhadap akurasi pengukuran. Suatu butir disebut bias jika didapati bahwa individu dengan karakteristik tertentu lebih diuntungkan dalam menjawab soal dibanding individu dengan karakteristik lain. Untuk soal nomor 30, dari 6 orang siswa perempuan hanya 1 orang yang menjawab benar. *Item-item* yang teridentifikasi *DIF* ($p < 0,05$) disarankan untuk di *review* ulang dan jika dirasa perlu direvisi atau diganti.

4. Simpulan

Dari pembahasan diatas dapat dikemukakan beberapa simpulan yaitu sebagai berikut: *pertama*, telah didapat hasil kajian tentang kualitas butir soal mata pelajaran Fisika pada *USBN* SMK tahun 2018 dengan menggunakan pemodelan

Rasch. Hasil analisis yang telah diteliti yaitu *reliability*nya baik person maupun item serta tingkat kesukaran soal. *Person reliability*nya lemah, namun *item reliability* bagus. *Kedua*, berdasarkan kriteria item yang tidak sesuai (*outliers* atau *misfits*), keseluruhan soal adalah fit. *Ketiga*, unidimensionalitas instrumen terpenuhi dengan nilai *raw varians* bagus untuk kelompok soal Lots dan cukup untuk kelompok soal Hots. *Keempat*, terdeteksi ada satu soal yang terjangkiti *DIF* yaitu soal nomor 30.

Hasil analisis ini tidak mutlak kebenarannya, karena setiap instrumen yang diujikan dalam setiap evaluasi dipengaruhi oleh banyak faktor. Bisa faktor dari instrument itu sendiri maupun dari segi responden (peserta didik) yang mengikuti ujian. Disarankan kegiatan menganalisis seperti ini dilakukan oleh guru untuk memetakan kemampuan siswanya setelah mengikuti pembelajaran di sekolah.

Daftar Pustaka

- Chiang, W.W. 2015. Ninth grade student self assessment in science: a Rasch analysis approach. *Procedia-sosial and behavioural science*: 200-210.
- Hofer, S.I et al. 2017. The test of basic Mechanics Conceptual understanding (bMCU): Using Rasch analysis to develop and evaluate an efficient multiple choice test on Newton's mechanics. *International journal of STEM education*. 4(18):1-20
- Kusuma, RT, Siahaan, SM & Andriani, N. 2019. Guided Inquiry model effect on student learning outcomes in Static Fluid. *Journal of Physics Conference Series*.1166: 1-6
- Marjiasih, K dan Wahyuni, S. 2014. Analisis kemampuan peserta didik dengan model Rasch. *Seminar nasional evaluasi pendidikan*: 94-102.
- Nitko, A.J dan Brookhart, S.M. 2007. *Educational Asesment of Students*. Ohio: Pearson.
- Palimbong, J., Mugasam & Allo, AYT. 2018. Item Analysis using Rasch Model in Semester Final Exam Evaluation Study Subject in Physics Class X TKJ SMK Negeri 2 Manokwari. *Kasuari: Physics Education Journal*.1 (1): 43-51
- Putra, A., Masril, M & Yurnetti, Y. 2018. Planning model of Physics Learning in Senior High School to Develop Problem Solving Creativity Based on National Standard of Education. *IOP Conference Series*. 335: 1-12
- Putri, FS & Istiyono, E. 2017. The Development of Performance Assesment of Stem Based Critical Thinking Skill in the High School Physics Lesson. *International journal of Environmental and Science Education*. 12(5): 1270-1279
- Rosana, D dan Sukardiyono. 2015. Analisis butir dan identifikasi ketidakwaajaran skor ujian akhir sekolah untuk standarisasi penilaian. *Jurnal kependidikan*. 45(2): 130-141.
- Safitri, AN, Sari, R & Wahyuni, S. 2017. The Influences of Mathematics Ability toward Physics Learning in Senior High School Based on an Autentic Assesment System. *International journal of learning and teaching*, 3(1): 11-14
- Solekhah, FM, Maharta, N & Suarna, W. 2018. Pengembangan Instrumen Tes Kemampuan berpikir tingkat tinggi pada materi Hukum Newton tentang Gerak. *Journal of Physics and Science Learning*. 2(1): 17-26
- Sumintono, B dan Widhiarso, W. 2014. *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Cimahi: trim Komunikata.
- Sumintono, B dan Widhiarso, W. 2015. *Aplikasi Pemodelan Rasch pada assessment Pendidikan*. Cimahi: Trim Komunikata.
- Widowati, T., Aminah, NS & Carli. 2016. Pengembangan Instrumen Penilaian Otentik Berbasis Scientific Literacy pada Pembelajaran Fisika di SMA sebagai Implementasi Kurikulum 2013. *Jurnal Inkuiri*. 5(2): 8-19