

## PENGEMBANGAN BUTIR SOAL TES TINGKAT MAKROSKOPIS, MIKROSKOPIS, DAN SIMBOLIS UNTUK IDENTIFIKASI PROFIL KEMAMPUAN ANALISIS SISWA

Feby Kristiansi✉, Endang Susilaningsih, Woro Sumarni, dan Sigit Priatmoko

Jurusan Kimia FMIPA Universitas Negeri Semarang

Gedung D6 Kampus Sekaran Gunungpati Telp. 8508112 Semarang 50229

### Info Artikel

Diterima : Jan 2022  
Disetujui : Feb 2022  
Dipublikasikan : Apr 2022

*Keywords: multilevel of representation, rate of reaction, analytical ability.*

Kata Kunci: Representasi MMS, laju reaksi, kemampuan analisis

### Abstrak

Penerapan kurikulum 2013 menuntut siswa memiliki kemampuan Higher Order Thinking (HOT). Kemampuan analisis merupakan salah satu indikator HOT yang dapat diidentifikasi dengan tes diagnostik. Penelitian ini bertujuan mengembangkan butir soal tes diagnostik Laju Reaksi tingkat Makroskopis, Mikroskopis, dan Simbolis (MMS) untuk identifikasi kemampuan analisis siswa. Metode penelitian ini adalah Research and Development (R&D), dengan desain model 4D. Subjek penelitiannya siswa SMA Negeri 2 Bantul. Instrumen pengumpul data meliputi lembar angket validasi, butir soal tes, dan lembar angket responden. Teknik Analisis data dimulai dari Analisis data validasi, analisis kualitas butir soal, dan analisis lembar angket responden. Hasil penelitian menunjukkan butir soal sangat valid dengan rerata skor validator 69,67 dari skor total 80 yang dapat dinyatakan layak digunakan. Analisis butir soal menggunakan ICT diperoleh reliabilitas 0,725, menggunakan model Rasch diperoleh person reliability 0,73 dan Item reliability 0,94. Hasil analisis lembar angket responden memiliki reliabilitas 0,848 dan menunjukkan respon positif dengan skor 40 dari skor total 48. Persen kepraktisan butir soal tes sebesar 85,8%. Profil kemampuan analisis siswa sebesar 56,62 % termasuk dalam kategori sedang. Simpulan penelitian telah diperoleh butir soal tes tingkat MMS yang teruji kelayakannya, validitas butir dan kepraktisannya yang mendapat respon positif.

### Abstract

*Implementation of the 2013 curriculum requires students to have Higher Order Thinking (HOT) abilities. Analytical ability is one of the HOT indicators that can be identified by diagnostic tests. This study aims to develop the test items for the Macroscopic, Microscopic, and Symbolic Reaction Rate (MMS) diagnostic test items to identify students' analytical skills. This research method is Research and Development (R&D), with a 4D model design. The research subjects are students of SMA Negeri 2 Bantul. Data collection instruments include validation questionnaire sheets, test items, and respondent questionnaire sheets. The data analysis technique starts from validation data analysis, item quality analysis, and respondent questionnaire sheet analysis. The results showed that the items were very valid with an average validator score of 69.67 out of a total score of 80 which could be declared suitable for use. Analysis of items using ICT obtained reliability of 0.725, using the Rasch model obtained person reliability of 0.73 and item reliability of 0.94. The results of the analysis of the respondent's questionnaire sheet have a reliability of 0.848 and show a positive response with a score of 40 out of a total score of 48. The practicality of the test items is 85.8%. The student's analytical ability profile of 56.62% is included in the medium category. The conclusion of the study was that MMS level test items were tested for their feasibility, item validity and practicality which received a positive response.*

Alamat korespondensi :

✉ E-mail: febykristiansi@gmail.com

© 2019 Universitas Negeri Semarang

ISSN NO 2252-6609

## Pendahuluan

Penerapan kurikulum 2013 menjadi salah satu jalan yang ditempuh pemerintah dalam suatu pembelajaran aktif, dimana proses pembelajaran aktif meliputi bertanya, mengamati, mengkomunikasikan dan menghubungkan konsep. Penerapan kurikulum 2013 dalam pembelajaran tersebut dilakukan dengan menggunakan pendekatan ilmiah (*scientific approach*). Pendekatan ilmiah merupakan pembelajaran yang dilakukan dengan mengadopsi langkah saintis melalui metode ilmiah dalam membangun pengetahuan (Retno, 2017). Pendekatan ilmiah yang diterapkan dalam pembelajaran ini melahirkan sistem evaluasi yang autentik. Guru harus mampu membuat progres ketercapaian akademik siswa dalam bidang kognitif yang merupakan hal utama dalam sebuah pembelajaran (Harsh, 2017).

Kurikulum 2013 tersebut juga menuntut siswa untuk memiliki keterampilan berpikir pada level yang lebih tinggi atau disebut *high order thinking* (HOT). Maka dari itu sesuai dengan tuntutan kurikulum 2013, siswa diharuskan tidak hanya bisa mengingat materi saja tetapi juga harus bisa menganalisis sesuai pernyataan Anderson (2001) yang menyatakan bahwa indikator untuk mengukur keterampilan berpikir tingkat tinggi meliputi menganalisis, mengevaluasi, dan menciptakan. Tujuan utama dari HOT adalah bagaimana meningkatkan kemampuan berpikir siswa pada level yang lebih tinggi (Dinni, 2018). Saputra (2016) menjelaskan bahwa HOT merupakan suatu proses berpikir siswa dalam level kognitif yang lebih tinggi yang dikembangkan dari berbagai konsep dan metode kognitif. Hal ini juga diperkuat dengan pernyataan Lailly (2015) bahwa HOT meliputi aspek kemampuan berpikir kritis, kemampuan berpikir kreatif, dan kemampuan memecahkan masalah/ analisis.

Kemampuan analisis dapat diidentifikasi menggunakan instrumen tes diagnostik. Instrumen tes tersebut dapat menjadi alternatif untuk mengidentifikasi kemampuan berpikir tingkat tinggi terutama kemampuan analisis menggunakan tipe multiple choice yang dimodifikasi. Modifikasi instrumen tes tipe multiple choice tersebut dikenal dengan sebutan *two tier multiple choice* (pilihan ganda bertingkat) yang dikembangkan oleh Treagus. Instrumen tes diagnostik two tier multiple choice (pilihan ganda bertingkat dua) telah berhasil meneliti dan memberikan hasil yang

baik dalam evaluasi kemampuan berpikir tingkat tinggi (Liana, 2018). Selain itu juga dapat mengidentifikasi miskonsepsi yang dialami siswa dalam belajar suatu materi khususnya ilmu kimia.

Ilmu kimia pada hakikatnya akan mudah dipahami apabila siswa mampu merepresentasikan ilmu kimia ke dalam tiga tingkat representasi. Ketiga tingkat representasi tersebut yaitu makroskopis, mikroskopis atau mikroskopis, dan simbolis (Cheng, 2009). Representasi makroskopis adalah representasi yang diperoleh melalui pengamatan secara nyata terhadap fenomena yang dapat dilihat baik secara langsung atau tak langsung dan dipersepsi oleh pancaindra seperti pembentukan gas, perubahan suhu, perubahan warna dalam reaksi kimia, dan terbentuknya endapan. Representasi mikroskopis adalah representasi yang memberikan penjelasan tingkat partikulat atau dapat digunakan untuk menjelaskan peristiwa dalam reaksi kimia dimana materi digambarkan sebagai molekul, atom, dan ion. Sedangkan representasi simbolis adalah representasi yang melibatkan penggunaan gambar atau simbol kimia, rumus senyawa, diagram, persamaan kimia, atau tulisan.

Penelitian di lapangan menyebutkan bahwa salah satu penghambat pembelajaran di kelas adalah kurang mampunya siswa dalam menghubungkan ketiga tingkat representasi tersebut (Safitri, 2019). Sehingga tujuan dari penelitian ini adalah mengembangkan butir soal tes tingkat makroskopis, mikroskopis, dan simbolis untuk identifikasi profil kemampuan analisis siswa pada materi Laju Reaksi dengan analisisnya menggunakan model Rasch. Pengembangan butir soal tes tersebut dimaksudkan untuk membantu guru dalam mengetahui profil kemampuan analisis siswa terhadap materi Laju Reaksi yang telah disampaikan. Sehingga dapat digunakan dalam perbaikan dalam proses pembelajaran dan remediasi sebagai langkah meningkatkan hasil belajar siswa.

## Metode Penelitian

Penelitian dilakukan di SMA Negeri 2 Bantul pada 5 Februari 2021 hingga 5 April 2021. Jenis penelitian yang digunakan dalam penelitian ini adalah penelitian pengembangan *Research and Development* (R&D) menggunakan metode 4D yang terdiri atas empat tahap yaitu *define* (pendefinisian), *design* (perancangan), *development* (pengembangan), dan *disseminate*

(penyebaran). Subjek dalam penelitian ini adalah 74 siswa kelas XI MIPA 4 hingga MIPA 6. Teknik pengumpulan dilakukan dengan observasi melalui sebuah wawancara kepada guru kimia, tes diagnostik kemampuan analisis melalui uji coba skala kecil hingga implementasi, dan pengisian angket responden siswa dan angket responden guru untuk mengetahui kepraktisan instrumen.

Tahap define dilakukan dengan mengumpulkan informasi-informasi dan observasi melalui wawancara guru kimia di SMA Negeri 2 Bantul. Tahap design dimulai dari langkah-langkah membuat lembar angket validasi, butir soal tingkat MMS untuk identifikasi kemampuan analisis siswa pada materi Laju Reaksi, dan lembar angket respon. Pada tahap development produk berupa butir soal tes diagnostik *two-tier multiple choice* tingkat MMS pada materi Laju Reaksi untuk identifikasi kemampuan analisis siswa ini dilakukan beberapa validasi yaitu validasi isi oleh ahli soal, validasi butir dengan analisis *Item Classical Theory* (ICT), dan validasi konstruk dengan *Item Response Theory* (IRT) model Rasch.

Instrumen tes diagnostik kemampuan analisis divalidasi terlebih dahulu oleh ahli sebelum diujicobakan. Setelah dinyatakan layak untuk diujicobakan, instrumen dapat digunakan pada uji coba skala kecil, skala besar, dan implementasi. Pada setiap tahap uji coba dilakukan perbaikan instrumen terlebih dahulu dengan mempertimbangkan hasil analisis ICT dan IRT. Analisis IRT model Rasch dipilih karena analisis ini memiliki keakuratan yang tinggi dan dapat menganalisis tidak hanya pada item, melainkan dapat menganalisis person Terakhir tahap disseminate merupakan tahap dimana setelah produk layak digunakan melalui beberapa uji, kemudian menginformasikan instrumen tersebut secara publik dengan tujuan dapat dijadikan rujukan pada penelitian-penelitian selanjutnya.

### Pembahasan

Hasil dari penelitian tahap define

berupa studi literatur dimana didapatkan hasil bahwa kemampuan berpikir tingkat tinggi sangat penting dimiliki oleh siswa pada abad ke-21 ini. Proses pembelajaran pada abad 21 perlu mengintegrasikan content knowledge yang dapat dilakukan dengan kegiatan pembelajaran yang merujuk pada proses menganalisis, mengevaluasi, dan mencipta. Kemampuan analisis siswa dapat dilatih dengan kegiatan yang bersifat menganalisis. Instrumen tes yang dapat menjadi alternatif untuk mengidentifikasi kemampuan berpikir tingkat tinggi terutama kemampuan analisis siswa salah satunya adalah menggunakan soal tipe *multiple choice yang dimodifikasi*. *Modifikasi instrumen tes tipe multiple choice tersebut dikenal dengan sebutan two tier multiple choice* (pihan ganda bertingkat) yang dikembangkan oleh Treagus pada tahun 1988 (Liana et al., 2018).

Setelah dilakukan penelitian pada tahap define dan didapatkan berbagai informasi yang dibutuhkan, tahap selanjutnya yaitu design perancangan butir soal *two-tier multiple choice* tingkat makroskopis, mikroskopis, dan simbolis yang memuat indikator kemampuan analisis. Tahap design ini dihasilkan (1) lembar angket validasi, (2) draf butir soal, (3) lembar angket respon guru dan siswa. Butir soal tes *two-tier multiple choice* tingkat makroskopis, mikroskopis, dan simbolis untuk identifikasi kemampuan analisis siswa disusun berdasarkan indikator dan kisi-kisi yang telah disusun sebelumnya. Butir soal dibuat sedemikian rupa agar menimbulkan proses berpikir analisis dari suatu permasalahan yang disajikan. Instrumen tes terdiri atas 25 butir soal pilihan ganda bertingkat materi Laju Reaksi. Tahap selanjutnya yaitu pengembangan (development), berikut hasil analisis dari tahap development.

Validitas butir soal dihitung menggunakan rumus rpbis dengan bantuan Microsoft Excel. Indeks korelasi point biserial (rpbis) yang diperoleh dari hasil perhitungan kemudian dianalisis. Soal dikatakan valid apabila nilai rpbis berada pada angka antara 0.30-0.70. Hasil soal yang tidak valid

Tabel 1. Butir soal yang tidak valid

Nomor Soal Tidak Valid	r <sub>pbis</sub>	t <sub>hitung</sub>	t <sub>tabel</sub> (α=0,05)	Validitas
8	0.105010851	0.896000537	1.666294	Tidak Valid
11	0.816746089	0.452333964	1.666294	Tidak Valid
12	0.183071676	1.580119381	1.666294	Tidak Valid
18	-0.006823811	-0.057903305	1.666294	Tidak Valid
21	0.077287589	0.657774451	1.666294	Tidak Valid
23	-0.017605432	-0.149410204	1.666294	Tidak Valid

berdasarkan nilai rpbis kemudian dianalisis lagi dengan menghitung t tabel dan melihat r tabel pada taraf signifikansi 5% yang disesuaikan dengan jumlah subjek yang diteliti. Soal dikatakan valid apabila  $t_{hitung} > t_{tabel}$  dengan  $dk=n-2$ . Rekap butir soal yang tidak valid dapat dilihat pada Tabel 1.

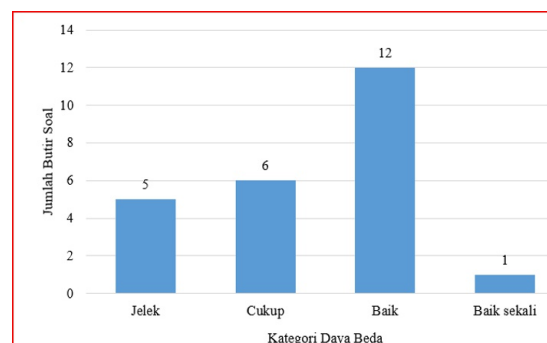
Nilai reliabilitas butir soal tes diagnostik *two-tier multiple Choice* tingkat MMS pada materi Laju Reaksi dalam penelitian ini dihitung menggunakan rumus Kurder Richardson (KR-20) dengan bantuan Microsoft Excel. Instrumen tes dikatakan reliabel apabila memiliki nilai  $r_{11} \geq 0.7$ . Berdasarkan perhitungan, reliabilitas instrumen tes sebesar 0.7259 yang berarti instrumen tes reliabel karena  $r_{11} \geq 0.7$  (Arikunto, 2014).

Hasil analisis tingkat kesukaran butir soal menunjukkan bahwa instrument tes memiliki jumlah butir soal dengan kategori sedang yang lebih banyak dibandingkan kategori mudah dan sukar. Instrumen tes yang baik adalah instrumen dengan butir soal yang tingkat kesukarannya tidak terlalu mudah dan tidak terlalu sukar. Melainkan memiliki taraf kesukaran yang sedang, sehingga siswa dengan abilitas rendah tidak kesulitan mengerjakan dan siswa dengan abilitas tinggi tidak terlalu mudah mengerjakan. Berikut adalah sebaran tingkat kesulitan butir soal yang dapat dilihat pada Tabel 2.

Analisis daya beda ini bertujuan untuk mengetahui sejauh mana butir soal tes dapat membedakan mana siswa yang menguasai materi laju reaksi dan mana siswa yang belum menguasai materi. Berdasarkan hasil daya beda pada penelitian ini menunjukkan bahwa butir soal baik, hal ini dibuktikan jumlah soal dengan kategori cukup, baik, dan baik sekali memiliki jumlah yang lebih banyak dibandingkan soal dengan kategori jelek. Dengan begitu soal sudah mampu membedakan dengan baik mana siswa yang menguasai materi dan yang tidak

Tabel 2. Rekap tingkat kesulitan butir soal

Kategori	Keterangan
	Uji Implementasi
Sukar	9, 10, 11, 12, 18, 19, 20, 23, 24
Sedang	1, 2, 3, 4, 7, 8, 13, 14, 15, 16, 17, 21, 22, 25
Mudah	5, 6



Gambar 1. Sebaran jumlah daya beda instrumen soal

menguasai. Berikut sebaran jumlah daya beda instrumen soal dapat dilihat pada Gambar 1.

Reliabilitas butir soal dalam analisis model Rasch dibagi menjadi person reliability dan item reliability. Person reliability merupakan nilai reliabilitas siswa dan *item reliability* merupakan nilai reliabilitas item soal. Reliabilitas dengan pemodelan Rasch dapat diketahui melalui nilai separasi individu (person separation) dan nilai separasi butir (item separation). Berikut hasil reliabilitas instrumen menggunakan analisis model Rasch disajikan pada Gambar 2.

Berdasarkan Gambar 2 di atas menunjukkan terdapat dua nilai pada setiap reliabilitas. Semakin kecil selisih antara dua nilai tersebut menunjukkan konsistensi siswa dalam menjawab soal tes yang cukup baik dan kualitas soal tergolong bagus. Reliabilitas item juga tergolong istimewa berdasarkan sebaran kategori nilai person reliability dan item

REAL RMSE	.54	TRUE SD	.81	SEPARATION	1.52	Person RELIABILITY	.70
MODEL RMSE	.50	TRUE SD	.84	SEPARATION	1.66	Person RELIABILITY	.73
S.E. OF Person MEAN = .11							
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00							
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .74 SEM = 2.07							
STANDARDIZED (50 ITEM) RELIABILITY = .85							
REAL RMSE	.32	TRUE SD	1.29	SEPARATION	4.05	Item RELIABILITY	.94
MODEL RMSE	.30	TRUE SD	1.29	SEPARATION	4.23	Item RELIABILITY	.95
S.E. OF Item MEAN = .27							

Gambar 2. Person reliability dan item reliability

reliability. Nilai alpha cronbach (reliabilitas) dalam analisis Rasch model ini menunjukkan nilai sebesar 0,74 yang termasuk dalam kategori bagus.

Tingkat kesesuaian butir soal atau item fit order dalam analisis model Rasch ini merupakan perhitungan yang digunakan untuk menentukan validitas butir soal. Butir soal yang dikatakan fit (cocok) berarti butir soal tersebut berperilaku secara normal dengan yang diharapkan oleh model, dalam hal ini adalah menganalisis. Kriteria yang digunakan agar butir soal dinyatakan fit (cocok) yaitu apabila diterima pada ketiga nilai berikut: (1) Nilai Outfit mean square (MNSQ) yang diterima:  $0,5 < \text{MNSQ} < 1,5$ ; (2) Nilai Outfit Z-standar (ZSTD) yang diterima:  $-2,0 < \text{ZSTD} < +2,0$ ; (3) Nilai Point Measure Correlation (Pt Mean Corr):  $0,4 < \text{Pt Measure Corr} < 0,85$ . Berikut adalah cuplikan hasil item fit order yang kurang baik pada penelitian ini disajikan pada Gambar 3.

Butir soal yang tidak fit pada penelitian ini hanya butir soal nomor 18. Butir soal tersebut terlihat tidak memenuhi ketiga kriteria nilai yang diterima pada analisis item fit order model Rasch yaitu nilai MNSQ, ZSTD, dan pt measure corr. Hal tersebut menunjukkan bahwa butir soal nomor 18 kurang berperilaku normal terhadap apa yang diharapkan oleh model yaitu soal untuk identifikasi kemampuan analisis. Setelah dibandingkan dengan analisis klasik ICT, soal nomor 18 memang termasuk dalam soal yang tidak valid.

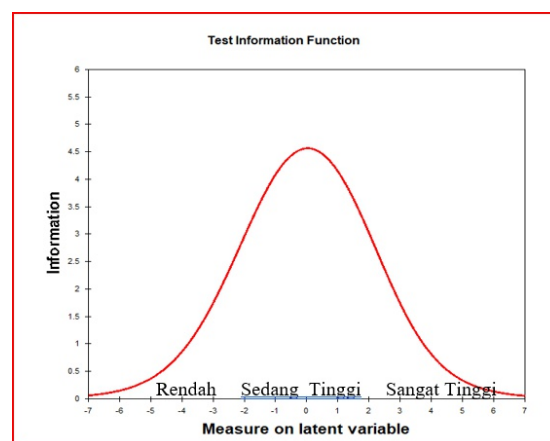
Item measure adalah analisis butir yang digunakan untuk mengetahui tingkat kesulitan butir soal. Tingkat kesulitan butir soal dapat diketahui melalui nilai logit butir yang terdapat pada kolom measure. Analisis model Rasch menunjukkan bahwa butir soal nomor 12, 18, 19, 23, dan 24 tergolong dalam soal yang sangat sulit dengan nilai logit diatas +SD. Sedangkan soal dengan kategori mudah yaitu butir soal nomor 3, 5, 6, dan 8 dengan nilai logit lebih kecil dari -SD. Berdasarkan hasil analisis tersebut, instrumen tes termasuk baik karena

jumlah soal dengan item measure dengan kategori sedang lebih banyak dibandingkan kategori sulit dan mudah.

Person measure atau disebut juga tingkat abilitas individu ini berfungsi untuk menganalisis tingkat kemampuan siswa dalam menjawab suatu soal. Secara keseluruhan, informasi yang tertera dalam person measure sama dengan informasi yang ada pada item measure. Analisis model Rasch menunjukkan bahwa 9.4% siswa memiliki tingkat abilitas yang sangat tinggi, 12% memiliki tingkat abilitas tinggi, 42% memiliki tingkat abilitas sedang, 57% memiliki tingkat abilitas yang rendah dengan nilai logit lebih kecil dari -SD.

Setiap pengukuran akan menghasilkan informasi mengenai hasil pengukuran. Informasi pengukuran ini bergantung antara hubungan Item dengan individu. Informasi yang didapatkan menunjukkan abilitas individu dalam mengerjakan soal tes. Pada Gambar 4 Menunjukkan sumbu X menunjukkan nilai abilitas individu (peserta tes), dan sumbu Y menunjukkan besarnya informasi fungsi.

Berdasarkan grafik tersebut terlihat sebaran nilai abilitas individu sedang memiliki informasi fungsi pengukuran yang paling tinggi. Sedangkan pada abilitas individu sangat tinggi, tinggi, dan rendah justru memiliki informasi fungsi pengukuran yang lebih rendah dibandingkan dengan abilitas individu sedang.



Gambar 4. Grafik fungsi pengukuran

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
18	9	74	1.81	.38	1.28	1.09	2.46	2.61	A-.02	.33	85.1	88.0	S18
23	9	74	1.81	.38	1.30	1.16	1.98	1.95	B .00	.33	85.1	88.0	S23
12	11	74	1.55	.35	1.12	.57	1.83	1.96	C .17	.35	85.1	85.7	S12
11	17	74	.92	.30	1.37	2.09	1.51	1.84	D .05	.39	70.3	79.5	S11

Gambar 3. Item fit order

Hal ini menunjukkan bahwa butir soal tes diagnostik tingkat MMS untuk identifikasi kemampuan analisis siswa ini memberikan informasi yang baik atau optimal jika diujikan kepada siswa dengan tingkat abilitas sedang pada skala yang besar. Selain itu, informasi fungsi pengukuran juga mampu membaca tingkat reliabilitas pengukuran. Semakin tinggi puncak, maka akan semakin tinggi nilai reliabilitas pengukurannya.

Tingkat kesesuaian individu (Person Fit Order) ini dapat mendeteksi jika terdapat individu yang pola responnya tidak sesuai (Sumintono, 2015). Pola repon tidak sesuai yang dimaksud adalah adanya ketidaksesuaian jawaban yang diberikan siswa yg disebabkan oleh respon siswa yang diduga menjawab secara terburu-buru, untung-untungan, bahkan menyontek. Kriteria penilaian yang digunakan dalam analisis person fit ini tidak berbeda dengan kriteria item fit, yaitu MNSQ, ZSTD, dan pt measure corr. Berikut hasil analisis Person Fit Order disajikan pada Gambar 5.

Berdasarkan hasil analisis person fit order yang disajikan pada Gambar 4.20 menunjukkan bahwa siswa yang tidak fit adalah siswa 60P, 21L, 69P, 71P, 01P, dan 68P tidak memenuhi ketiga kriteria nilai yang diterima yaitu nilai *outfit mean square* (MNSQ), outfit Z-standar (ZSTD), dan nilai point measure correlation (Pt Mean Corr). Sedangkan siswa beberapa nomor dengan tanda merah selain yang sudah disebutkan tidak memenuhi pada kedua kriteria nilai yang diterima yaitu *outfit mean square* (MNSQ) dan nilai point measure correlation (Pt Mean Corr). Nilai *point measure correlation* yang negatif mengindikasikan adanya pola respon di luar kebiasaan.

Berdasarkan Gambar 5 di atas, siswa 60P tersebut memiliki nilai logit abilitas sebesar -2,17 dimana nilai logit abilitas tersebut termasuk dalam kategori rendah. Siswa 40P

dapat menjawab 4 butir soal yang benar yaitu butir 4, 12, 22, dan 23. Empat butir tersebut memiliki nilai logit sebesar -1,06 logit, 1,55 logit, -0,14 logit, dan 1,81 logit. Logit item tersebut menunjukkan nilai yang lebih besar dibandingkan dengan logit abilitas siswa itu sendiri, kecuali butir nomor 4. Bahkan siswa dengan abilitas sedang dapat menjawab dengan benar soal nomor 23 yang termasuk dalam soal sangat sulit. Hal ini menunjukkan atau sangat dimungkinkan siswa 60P menjawab secara untung-untungan atau menebak, sehingga menyebabkan pola respon siswa menjadi tidak fit. Begitu pula pada siswa yang tidak fit lainnya. Hal tersebut disebabkan adanya faktor untung-untungan atau menebak.

Selain hasil analisis person fit, informasi berkaitan dengan respon siswa juga dapat dijelaskan secara rinci melalui output scalogram dari software ministep. Melalui matriks Guttman ini akan diketahui siswa yang melakukan kecurangan saat mengerjakan tes. Berikut scalogram pada uji coba implementasi dapat dilihat pada Gambar 6.

Berdasarkan Gambar 6 jika kita periksa pada siswa dengan kode 53P dan 55P, 04P hingga 23L, 47P dan 50P, kemudian 61L-63P memiliki pola jawaban yang sama. Pola jawaban tersebut yaitu menjawab dengan jumlah benar yang sama dan pada nomor soal yang benar juga sama. Hal ini menunjukkan bahwa siswa tersebut bekerja sama dalam mengisi jawaban tes. Analisis ini sangat penting dan bermanfaat bagi pendidik untuk menilai kemampuan setiap siswa dengan akurat.

Profil kemampuan analisis siswa pada penelitian ini dihitung berdasarkan jawaban benar pada kedua pilihan jawaban yaitu tier-1 dan tier-2. Mengapa digunakan jawaban benar pada kedua tier? Hal ini berkaitan dengan model yang diinginkan yaitu profil kemampuan analisis, dimana kemampuan analisis berarti

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL		INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		Person
				S. E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
60	4	25	-2.17	.61	1.66	1.66	4.48	2.64	A-.18	.41	76.0	85.4	60P	
21	15	25	.57	.47	1.77	3.03	3.36	4.22	B-.13	.50	56.0	73.2	21L	
69	7	25	-1.26	.51	1.73	2.49	3.36	3.18	C-.10	.48	64.0	77.4	69P	
59	4	25	-2.17	.61	1.42	1.16	2.88	1.82	D-.08	.41	76.0	85.4	59P	
18	5	25	-1.83	.56	1.71	2.01	2.73	1.98	E-.09	.44	72.0	82.7	18P	
71	14	25	.34	.47	1.35	1.58	2.42	3.18	F-.19	.51	64.0	72.8	71P	
1	9	25	-.77	.48	1.81	3.03	2.26	2.58	G-.05	.50	56.0	74.6	01P	
68	9	25	-.77	.48	1.87	3.20	2.25	2.57	H-.08	.50	48.0	74.6	68P	
38	5	25	-1.83	.56	1.03	.21	2.07	1.44	I-.33	.44	88.0	82.7	38L	
24	6	25	-1.53	.53	1.34	1.22	1.86	1.39	J-.20	.46	72.0	79.7	24P	
74	4	25	-2.17	.61	1.38	1.07	1.71	.99	K-.12	.41	84.0	85.4	74L	
72	10	25	-.54	.48	1.11	.56	1.70	1.79	L-.38	.51	76.0	74.0	72P	
26	6	25	-1.53	.53	.97	-.03	1.64	1.13	M-.42	.46	88.0	79.7	26P	

Gambar 5. Person fit order

53	+1111111110111111110000010	53P
55	+1111111110111111110000010	55P
4	+111111111100000001000000	04P
8	+111111111100000001000000	08L
9	+111111111100000001000000	09P
11	+111111111100000001000000	11P
12	+111111111100000001000000	12P
15	+111111111100000001000000	15L
22	+111111111100000001000000	22P
23	+111111111100000001000000	23L
47	+1110000000011101100000100	47P
50	+1110000000011101100000100	50P
61	+111111110100000000000000	61L
62	+111111110100000000000000	62P
63	+111111110100000000000000	63P

Gambar 6. Scalogram

siswa mampu mengorganisasikan, membedakan, serta mengatribusi setiap informasi yang ada. Soal dirancang saling berkaitan antara pilihan jawaban tier-1 dengan pilihan jawaban tier-2. Sehingga siswa harus mampu mengaitkan antara pilihan jawaban yang dipilih pada tier-1 dengan pilihan jawaban yang dipilih pada tier-2. Dengan demikian kemampuan analisis dihitung berdasarkan jawaban benar pada kedua tier, karena apabila hanya benar pada salah satu tier, itu mengindikasikan adanya miskonsepsi.

Secara keseluruhan profil kemampuan analisis siswa pada uji implementasi menunjukkan persentase sebesar 56.62% yang termasuk dalam kategori sedang. Sedangkan persentase tertinggi pada indikator kemampuan analisis siswa terdapat pada indikator attributing yaitu sebesar 34.79% yang termasuk dalam kategori sedang.

Angket respon siswa ini menjadi dasar untuk perbaikan-perbaikan terhadap instrumen tes agar menjadi lebih baik. Selain respon siswa, respon guru juga diperlukan untuk perbaikan instrumen. Angket responden siswa terdiri dari 9 butir pernyataan berkaitan dengan hal yang dirasakan siswa setelah mengerjakan instrumen tes. Jumlah maksimal setiap siswa adalah 36. Berdasarkan hasil analisis angket menunjukkan responden siswa dan guru, siswa memberikan respon kurang positif yaitu banyak siswa yang memberikan nilai dua pada angket. Hal ini dikarenakan siswa masih merasa kesulitan dalam mengerjakan soal tes dan belum terbiasa dengan soal diagnostik. Sedangkan guru memberikan respon yang sangat positif yaitu dengan total skor 40 dari 48 poin maksimal. Instrumen tes diagnostik level makroskopis, mikroskopis, dan simbolis untuk identifikasi kemampuan analisis

siswa pada materi Laju Reaksi ini juga dinyatakan praktis digunakan. Hal tersebut berdasarkan persentase kepraktisan yang diperoleh dari angket responden siswa yang menyatakan soal tersebut praktis yaitu sebesar 86.48%.

### Simpulan

Pengembangan Butir Soal Tes Tingkat MMS pada Materi Laju Reaksi berhasil mengidentifikasi kemampuan analisis siswa secara baik dengan kualitas instrumen yang teruji secara akurat. Profil kemampuan analisis siswa secara keseluruhan yaitu sebesar 56.62%, sedangkan profil kemampuan analisis siswa berdasarkan indikator dibagi menjadi 3 yaitu differentiating, organizing, dan attributing. Persentase profil kemampuan analisis siswa pada setiap indikator yaitu: 14.05% pada indikator differentiating, 7.77% pada indikator organizing, dan 34.79% pada indikator attributing. Instrumen tes yang dikembangkan juga memiliki persen kepraktisan sebesar 86.48%. Dengan adanya hasil analisis ini diharapkan dapat menjadi evaluasi dalam pembelajaran khususnya pembelajaran di masa pandemi Covid-19.

### Ucapan Terima Kasih

Ucapan terima kasih disampaikan kepada kedua orang tua, dosen pembimbing, dosen penguji, guru pembimbing, siswa SMA Negeri 2 Bantul, serta seluruh pihak yang telah membantu dalam menyelesaikan penelitian ini.

### Daftar Pustaka

- Anderson, L., & Krathwohl, D. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman
- Cheng, M. M., & Gilbert, J. K. (2009). *Toward a Better Utiliation of Diagrams in Research into the Use of Representative Levels in Chemical Education* (Vol. 4). (J. Gilbert, & D. Treagust, Eds.) Scotland: Springer. doi:10.1007/978-1-4020-8872-8-4
- Dinni, H. N. (2018). HOTS (High Order Thinking Skills) dan kaitannya dengan kemampuan literasi matematika. PRISMA (Prosiding Seminar Nasional Matematika) (pp. 170-176). Semarang: Unnes.
- Harsh, J., Esteb, J. J., & Maltese, A. V. (2017). *Evaluating the Development of Chemistry Undergraduate Researchers' Scientific Thinking Skills Using Performance-Data: First Findings from the Performance Assessment of Undergraduate Research (PURE) Instrument*. *Chemistry Education Research and Practice*, 18(3), 472-485.

- Lailly, N., & Wisudawati, A. (2015). Analisis Soal Tipe HOTS Dalam Soal UN Kimia SMA Rayon B Tahun 2012/2013. *Integration and Interconnection Islam and Science*, 11(1), 27-39.
- Liana, N., Suana, W., Sesunan, F., & Abdurrahman. (2018). Pengembangan Soal Tes Berpikir Tingkat Tinggi Materi Fluida untuk SMA. *Journal of Komoda Science Education*, 1(1), 66-78.
- Retno, A. T., Saputro, S., & Ulfa, M. (2017). Kajian Aspek Literasi Sains pada Buku Ajar Kimia SMA Kelas XI di Kabupaten Brebes. *Seminar Nasional Pendidikan Sains* (pp. 112-123). Surakarta: Universitas Negeri Surakarta.
- Safitri, N. C., Nursa'adah, E., & Wijayanti, I. E. (2019). Analisis Multi Representasi Kimia Siswa pada Konsep Laju Reaksi. *EduChemia (Jurnal Kimia dan Pendidikan)*, 4(1), 1-12.
- Saputra, H. (2016). Pengembangan Mutu Pendidikan Menuju Era Global: Penguatan Mutu Pembelajaran dengan Penerapan HOTS (High Order Thinking Skills). Bandung: SMILE's Publishing.
- Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch: pada Assessment Pendidikan. Cimahi: Trim Komunikata.