

## Pengembangan Soal Tes Tertulis Berbasis Stem Dengan Pemodelan Rasch Di Sekolah Dasar

**Adinda Aulia Rahmat<sup>1</sup>, Ghullam Hamdu<sup>2</sup>, Epon Nur'aeni<sup>3</sup>, Dindin Abdul Muiz Lidinillah<sup>4</sup>**

Program Studi PGSD Universitas Pendidikan Indonesia Kampus Tasikmalaya <sup>1234</sup>  
adindaar98@student.upi.edu<sup>1</sup>, ghullamh2012@upi.edu<sup>2</sup>,  
nuraeni@upi.edu<sup>3</sup>, dindin\_a\_muiz@upi.edu<sup>4</sup>

### Abstrak

Penelitian ini bertujuan mengembangkan soal tes tertulis berbasis STEM disertai analisisnya menggunakan pemodelan Rasch. Penelitian dilakukan menggunakan metode focus group discussion (FGD). Rendahnya keterampilan siswa dalam berpikir tingkat tinggi diketahui berdasarkan hasil capaian Indonesia dalam ajang TIMSS dan PISA. Siswa di Indonesia tidak terbiasa dengan soal yang memiliki karakteristik HOTS. Selain itu, berdasarkan studi lapangan masih jarang ditemukan soal tes tertulis yang mengembangkan keterampilan berpikir tingkat tinggi. Sehingga, memunculkan kebutuhan atas pengembangan soal tes tertulis HOTS. Soal HOTS dikembangkan memperhatikan salah satu bagian dari keterampilan abad 21 yaitu berpikir kritis. Hasil penelitian berupa 10 butir soal tes tertulis berbasis STEM yang dikembangkan berdasarkan indikator keterampilan berpikir kritis pada "*Assessing 21st Century Skills for Teachers and Students*". Soal tes tertulis disertai analisis menggunakan pemodelan Rasch berbantuan aplikasi winstep Rasch versi 4.4.6 yang hasilnya dapat menjelaskan kualitas dari setiap soal berdasarkan tingkat kesulitan dan kesesuaian butir soal, serta tingkat abilitas dan tingkat kesesuaian siswa.

**Kata kunci:** Soal tes tertulis, HOTS, pemodelan Rasch, STEM.

## PENDAHULUAN

Kualiatas pendidikan Indonesia harus mampu bersaing secara global. Karena pada saat ini, pendidikan dituntut untuk memenuhi tuntutan keterampilan abad 21. Salah satu keterampilan abad 21 yang perlu dimiliki siswa yaitu learning and innovation skill (Trilling & Fadel, 2009). Learning and innovation skills menjadi fokus pembicaraan yang diangkat dalam kurikulum 2013. Pada ranah learning and innovation skills didalamnya diuraikan menjadi berpikir kritis dan pemecahan masalah, kreativitas dan inovasi, komunikasi dan kolaborasi. Berdasarkan uraian tersebut maka banyak orang lebih mengenalnya dengan ketrampilan 4C yaitu critical thinking and problem solving, creativity and innovation, communication, dan collaboration. Dengan keterampilan tersebut akan mengantarkan siswa untuk dapat bersaing secara global, bahkan mampu bersaing dengan kecerdasan buatan (artificial intelligent)(Ariyanti, Dantes, & Marhaeni, 2020; Arseni, Lasmawan, & Suarni, 2020).

Keterampilan abad 21 memiliki keterkaitan dengan keterampilan berpikir tingkat tinggi atau HOTS (High Order Thinking Skills). Hal tersebut karena keterampilan berpikir tingkat tinggi atau HOTS merupakan proses berpikir pada siswa berada dalam level kognitif yang lebih tinggi karena dikembangkan dari berbagai konsep dan taksonomi berpikir meliputi kemampuan pemecahan masalah, kemampuan berpikir kreatif, berpikir kritis, kemampuan berargumen, dan kemampuan mengambil keputusan (Saputra, 2016). Diantara ke empat keterampilan abad 21 dan yang terkait dengan HOTS tersebut, critical thinking atau keterampilan berpikir kritis menjadi dasar dari keempat keterampilan tersebut. Dengan berpikir kritis, maka menjadikan modal awal untuk mengembangkan keterampilan berpikir tingkat tinggi termasuk mengembangkan keterampilan abad 21 lainnya. Keterampilan berpikir kritis dapat dikembangkan berdasarkan indikator pada *"Assessing 21st Century Skills for Teachers and Students"* meliputi *analyzing arguments, claiming or proving; making conclusion or reasoning; judging or evaluating; making decisions or problem solving* (P21 Partnership for 21st Century Skills: <http://www.p21.org/our-work/p21-framework>). Indikator tersebut dapat dijadikan referensi untuk menilai siswa (Hamdu, Fuadi, Yulianto, & Akhirani, 2020). Sejatinya, keterampilan berpikir kritis sangat diperlukan sebagai keterampilan dasar

oleh siswa Indonesia. Namun, keterampilan berpikir kritis siswa di Indonesia terbilang sangat rendah. Hal ini dapat dilihat dari kemampuan siswa Indonesia dalam menyelesaikan soal-soal pada ajang TIMSS dan PISA yang bersaing secara internasional. Hasil capaian Indonesia pada TIMSS (Trends in International Mathematics and Science Study) tahun 2015 mendapat peringkat 44 dari 49 negara (Mullis, Martin, Foy, & Hooper, 2015). Begitupun dengan hasil capaian Indonesia pada ajang PISA pada tahun 2018 hanya mencapai pada level 1 dengan skor rata-rata 396 (OECD, 2019). Sejalan dengan hal tersebut, data dilapangan menunjukkan bahwa bentuk soal yang dikembangkan di SD saat ini cenderung merupakan soal-soal LOTS yang hanya berkisar pada ranah kognitif mengingat, dan guru di sekolah dasar pada umumnya telah mengetahui istilah kemampuan berpikir tingkat tinggi dalam ranah kognitif siswa, hanya saja dalam penerapannya guru jarang menggunakan soal jenis kemampuan berpikir tingkat tinggi (Akhirani, 2019; Fitriani, 2018).

Bagi siswa, keterampilan berpikir tingkat tinggi perlu dibiasakan dan dilatih. Salah satu bentuk aktivitas yang dapat memberikan pengalaman untuk mengembangkan keterampilan berpikir tingkat tinggi dan menunjang untuk mengembangkan keterampilan abad 21 adalah berlatih melalui soal tes tertulis. Soal tes tertulis tersebut dapat dikembangkan dengan berbasis STEM (Sains, Technology, Engineering dan Mathematic) sehingga dapat menggali kemampuan siswa untuk mengembangkan keterampilan berpikir pada tingkat tinggi. Pendekatan STEM relevan dengan pengembangan keterampilan abad 21, karena keterampilan yang dihasilkan dari pembelajaran STEM tumpang tindih dengan keterampilan yang dibutuhkan dalam pendidikan abad ke-21 (Hamdu et al., 2020; Zulfikar, 2019) Melalui pendekatan STEM, siswa akan memperoleh pengetahuan secara terintegrasi dari sains dan matematika, praktek penyelidikan ilmiah, teknologi dan rekayasa, dan keterampilan abad 21 (Lidinillah, Mulyana, Karlimah, & Hamdu, 2019). Dengan demikian, pendekatan STEM dapat digunakan sebagai alternatif untuk menunjang pengembangan keterampilan berpikir tingkat tinggi dan keterampilan abad 21 melalui soal tes tertulis.

Dalam konteks belajar mengajar, soal tes tertulis erat kaitannya dengan penilaian. Penilaian dilakukan untuk mengumpulkan informasi yang berhubungan dengan kemajuan belajar peserta didik berdasarkan kompetensi yang harus dikuasai

(Hosnan, 2014). Dalam hal ini, diperlukan soal tes yang berkualitas (Hamdu et al., 2020). Soal tes tersebut harus dapat memungkinkan untuk dianalisis kualitasnya. Salah satu pemodelan pengukuran yang dapat digunakan yaitu pemodelan Rasch. Pemodelan Rasch menyediakan teknik analisis psikometri yang dapat digunakan untuk mengembangkan item tes dan untuk menyajikan informasi yang relevan terkait dengan penilaian belajar siswa (Sumintono, 2018). Dengan menggunakan pemodelan Rasch akan menghasilkan informasi yang lebih tepat (Sumintono & Widhiarso, 2015). Dikatakan lebih tepat, karena informasi yang diberikan dapat lebih informatif dan relevan.

Merujuk pada informasi terkait dengan rendahnya keterampilan siswa dalam berpikir tingkat tinggi dan perlunya instrument tes yang mengembangkan keterampilan berpikir tingkat tinggi yang berkualitas maka timbul kebutuhan berupa instrument penilaian dalam bentuk soal tes tertulis yang berkualitas. Oleh karena itu, peneliti hendak melakukan penelitian dengan tujuan mengembangkan soal tes tertulis berbasis STEM dengan pemodelan Rasch di sekolah dasar.

#### METODE PENELITIAN

Penelitian ini menggunakan metode *focus group discussion* (FGD) dengan pendekatan kualitatif. Prosedur penelitian seperti pada gambar 1 (Glynn, Shanahan, & Duggan, 2015).



Gambar1. Proses FGD

Pada tahapan *conduct research* , soal tes yang dikembangkan diujikan kepada 15 siswa kelas 5 SD Negeri 7 Ciamis. Hasil tes yang diperoleh dianalisis menggunakan pemodelan rasch berbantuan aplikasi Winstep Rasch versi 4.4.6. Data yang

diperoleh, kemudian di deskripsikan sehingga diperoleh kualitas soal berdasarkan tingkat kesulitan dan kesesuaian, dan abilitas serta kesesuaian siswa.

## **HASIL PENELITIAN**

### ***Identify goal/objective***

Pada tahap ini dilakukan identifikasi dan analisis masalah, sehingga diketahui kebutuhan dan tujuan dari penelitian yang akan dilakukan. Berdasarkan hasil wawancara, diketahui bahwa soal tes untuk ulangan harian dan penilaian tengah semester (PTS) dibuat oleh guru kelas. Soal tersebut dibuat berdasarkan kompetensi yang telah diajarkan dan disesuaikan dengan kemampuan siswa. Guru juga tidak menganalisis soal atau hasil pengerjaan siswa secara lebih lanjut karena nilai langsung diperoleh berdasarkan jawaban benar siswa. Terkait dengan soal berbasis STEM belum ditemukan, karena pembelajaran dengan berbasis STEM pun belum pernah dilaksanakan. Kemudian, berdasarkan hasil studi dokumentasi diketahui bahwa keberadaan soal HOTS masih jarang ditemukan. Sehingga dapat diketahui, bahwa diperlukan soal tes tertulis yang mengembangkan keterampilan berpikir tingkat tinggi. Selain itu, perlunya analisis lebih lanjut sehingga diketahui informasi kualitas dari soal tersebut. Dengan demikian perlu dilakukan focus group discussion sehingga

### ***Identify question***

Identifikasi pertanyaan berkaitan dengan identify goal pada tahap pertama. Pada dasarnya FGD bertujuan untuk mengumpulkan data mengenai persepsi dan pandangan peserta, sehingga pertanyaan lebih bersifat *open-ended*. Namun, pertanyaan pertama secara umum meliputi: 1) Bagaimana bentuk dan pengolahan soal tes tertulis yang digunakan di sekolah dasar pada umumnya? 2) Bagaimana rancangan awal soal tes tertulis berbasis STEM untuk sekolah dasar? 3) Bagaimana uji coba soal tes tertulis berbasis STEM untuk sekolah dasar? 4) Bagaimana proses pengolahan soal tes tertulis berbasis STEM dengan menggunakan pemodelan rasch? 4) Bagaimana bentuk akhir rancangan soal tes tertulis dengan menggunakan pemodelan rasch pada pembelajaran STEM di sekolah dasar?

### ***Identify people***

Jumlah peserta yang terlibat yaitu sebanyak 7 orang yang telah memenuhi kriteria. Peserta merupakan tim perangkat pembelajaran STEM. Sedangkan, moderator yaitu

peneliti sendiri dan dibantu fasilitator yang ahli dalam perangkat pembelajaran, kajian soal tes tertulis, STEM, dan pemodelan rasch sehingga dapat mendukung jalannya proses FGD.

#### ***Select time, place/environment***

Waktu yang dilakukan untuk satu sesi FGD yaitu 120 menit. Waktu tersebut disesuaikan dengan kebutuhan peneliti. Tempat dilakukan FGD yaitu di UPI Kampus Tasikmalaya. Selain itu FGD dilakukan juga melalui aplikasi Google Meet.

#### ***Conduct research***

Sebelum pada proses inti FGD peneliti terlebih dahulu membuat rancangan produk dengan langkah-langkah sebagai berikut:

#### **Analisis materi pembelajaran**

Pada tahap ini ditentukan materi pembelajaran berdasarkan kompetensi dasar pada tema tokoh dan penemuan. Berdasarkan hasil kajian ditentukan tiga mata pelajaran yang akan digunakan yaitu IPA, Matematika dan Bahasa Indonesia. Kemudian, mata pelajaran terpilih diproyeksikan kedalam matriks STEM.

#### **Analisis indikator**

Dilakukan analisis indikator berdasarkan pemetaan kompetensi dasar sebelumnya. Indikator tersebut kemudian dikembangkan menjadi indikator soal.

#### **Analisis tujuan pembelajaran**

Analisis tujuan pembelajaran dilakukan sehingga diperoleh tujuan yang jelas dalam bentuk pernyataan yang menunjukkan keterampilan siswa yang diharapkan dapat dicapai sebagai hasil belajar.

**Rancangan pengembangan soal tes tertulis berbasis STEM**

Rancangan soal tes tertulis terdiri dari soal seperti pada gambar 2 dan kisi-kisinya seperti pada gambar 3. Soal dikembangkan berbasis STEM dengan memperhatikan

5. Kelompok 2 membuat *Electrical Tandem Roller* dengan lampu yang disusun secara berderet. Seluruh komponen sudah terpasang dengan baik. Namun salah satu lampunya rusak, sehingga menyebabkan seluruh lampu yang terpasang menjadi tidak menyala. Berdasarkan informasi tersebut, apa yang menyebabkan seluruh lampu menjadi tidak menyala sedangkan hanya salah satu lampu saja yang dinyatakan rusak?
  - a. Lampu dirangkai secara seri, sehingga ketika salah satu lampu rusak maka lampu yang lainnya menjadi tidak menyala karena arus yang mengalir harus melewati satu lampu terlebih dahulu sebelum dialirkan pada lampu selanjutnya.
  - b. Lampu dirangkai secara paralel, sehingga ketika salah satu lampu rusak maka lampu yang lainnya menjadi mati karena arus yang mengalir harus melewati satu lampu terlebih dahulu sebelum dialirkan pada lampu selanjutnya..
  - c. Lampu dirangkai secara seri, sehingga ketika salah satu lampu rusak maka lampu lainnya menjadi sama-sama rusak karena arus yang dialirkan dibagikan pada setiap lampu yang ada.
  - d. Lampu dirangkai secara seri, sehingga terjadi korsleting atau arus pendek karena arus listrik yang mengalir menjadi sangat besar dan dalam waktu yang sangat cepat.

Gambar 2. Soal tes

keterampilan abad 21 dan keterampilan berpikir tingkat tinggi. Pada rancangan awal, soal terdiri dari 8 soal pilihan ganda dan 3 soal uraian yang dikembangkan berbasis STEM dengan indikator keterampilan berpikir kritis dan kreatif.

Kompetensi Dasar	Indikator Pencapaian Kompetensi	Indikator soal	Materi	No. soal	Keterampilan berpikir	Indikator keterampilan berpikir
4.4 Melakukan percobaan rangkaian listrik sederhana secara seri dan paralel	4.4.1 Membuat rancangan rangkaian listrik sederhana secara seri dan paralel	Disajikan gambar dari <i>Electrical Tandem Roller</i> , siswa dapat menentukan rangkaian lampu pada gambar dengan tepat.	Listrik	5	Critical thinking	reasoning

Gambar 3. Kisi-kisi soal

Setelah rancangan selesai dibuat, selanjutnya dilakukan kegiatan inti dari *focus group discussion*. Berdasarkan FGD diperoleh hasil sebagai berikut:

*Focus group discussion* (FGD) ke 1

Diperoleh hasil FGD 1 yaitu penambahan soal uraian mengenai rangkaian listrik dan perbaikan pada dua butir soal hitungan yang kurang kontekstual.

*Focus group discussion* (FGD) ke 2

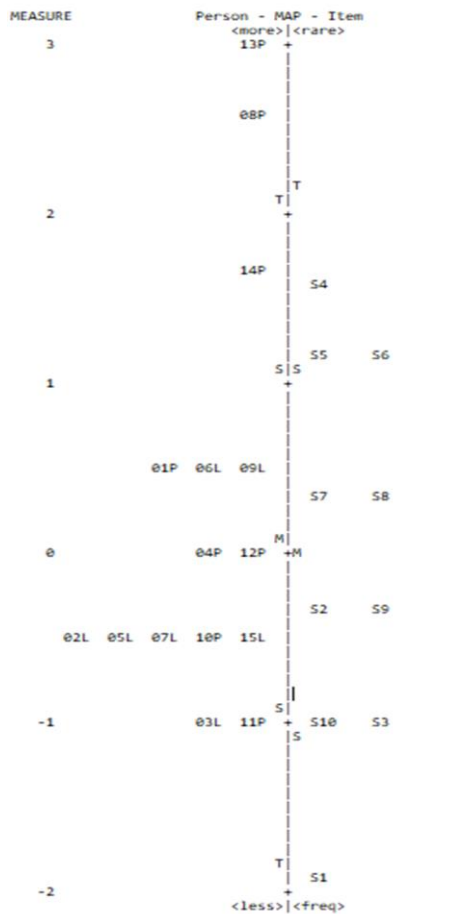
Berdasarkan hasil FGD ke 2, soal pilihan ganda ditambahkan sebanyak 2 butir soal yang dengan kriteria mengandung stimulus berupa gambar. Kemudian, terdapat perbaikan kalimat pada dua butir soal. Perbaikan kalimat tersebut dimaksudkan agar lebih efektif dan lebih mudah dipahami oleh siswa.

*Focus group discussion (FGD) ke 3*

Pada FGD ke 3 masih terdapat perbaikan pada satu soal, dengan harapan dapat lebih menstimulus keterampilan berpikir siswa. Selanjutnya diperoleh keputusan untuk analisis menggunakan pemodelan rasch, yaitu 10 butir soal pilihan ganda dengan keterampilan berpikir kritis yang dikembangkan berdasarkan indikator “*Assessing 21st Century Skills for Teachers and Students*”.

**Evaluate finding/data**

Dilakukan analisis soal tes menggunakan pemodelan rasch. Soal tes yang telah melalui proses FGD secara berulang dievaluasi melalui analisis pemodelan rasch berbantuan aplikasi winstep versi 4.4.6. Analisis tersebut dilakukan untuk mengetahui kualitas soal berdasarkan tingkat kesulitan dan kesesuaian soal. Selain itu dilakukan juga analisis tingkat ablitas dan kesesuaian siswa terhadap soal yang telah dibuat. Diperoleh data berdasarkan analisis peta wright (*person-item map*)



Gambar 4. Peta wright



seperti pada gambar 4.

Peta wright tersebut menggambarkan sebaran kemampuan siswa dalam mengerjakan soal tes tertulis berbasis STEM dan sebaran tingkat kesulitan soal dengan skala yang sama. Pada peta wright bagian kiri menggambarkan abilitas siswa, sehingga dapat diketahui terdapat satu orang siswa yang memiliki abilitas tinggi, yaitu 13P. Nilai *logit* dari siswa 13P adalah +3 *logit*. Selain itu, dapat diketahui juga siswa yang memiliki abilitas paling rendah, yaitu 03L dan 11P. Keduanya memiliki nilai -1 *logit*. Kemudian, pada bagian kanan peta wright menunjukkan distribusi tingkat kesulitan butir soal. Butir S4 memiliki tingkat kesukaran paling tinggi, dengan kata lain probabilitas semua siswa untuk mengerjakan soal dengan benar sangat kecil. Adapun butir S1 adalah soal yang paling mudah, dalam hal ini hampir semua siswa dapat mengerjakan soal tersebut dengan benar. Melalui peta wright dapat diketahui deskripsi secara umum mengenai abilitas siswa dan tingkat kesulitan butir soal. Dari deskripsi umum tersebut dapat memberikan interpretasi yang cukup jelas mengenai tingkat kesulitan butir soal. Analisis lebih lanjut dari peta Wright dapat dijelaskan menggunakan tabel yang lebih Analitis didasarkan pada distribusi kesulitan soal tes tertulis dan distribusi kemampuan siswa dari hasil tes tertulis.

Untuk dapat mengetahui tingkat kesulitan butir soal, maka dapat dikethui melalui *item measure* menggunakan analisis Rasch. Klasifikasi tingkat kesulitan item didasarkan pada kombinasi dari nilai standar deviasi (SD) dan nilai rata-rata *logit* (Sumintono & Widhiarso, 2015). Sehingga dapat dikethui tingkat kesulitan butir soal berdasarkan kelompok butir soal sukar yaitu lebih besar dari 1SD; kelompok butir soal sulit yaitu 0,0+1SD; kelompok butir soal mudah yaitu 0,0-1SD; dan kelompok butir soal sangat mudah yaitu -1SD. Berdasarkan klasifikasi tingkat kesulitan soal, maka dapat dianalisis tingkat kesulitan butir soal yang telah dibuat seperti pada gambar 5.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.		INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH	Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%		
4	4	15	1.59	.72	1.70	1.48	2.16	1.71	.17	.59	64.3	81.8	S4
5	5	15	1.13	.66	.61	-1.19	.49	-1.31	.79	.55	85.7	77.6	S5
6	5	15	1.13	.66	1.26	.82	1.28	.73	.40	.55	71.4	77.6	S6
7	7	15	.37	.59	.78	-.99	.71	-.97	.63	.48	78.6	68.5	S7
8	7	15	.37	.59	.78	-.99	.71	-.97	.63	.48	78.6	68.5	S8
2	9	15	-.30	.58	.84	-.88	.76	-.60	.51	.41	78.6	65.3	S2
9	9	15	-.30	.58	.83	-.94	.76	-.61	.52	.41	78.6	65.3	S9
3	11	15	-1.01	.62	1.19	.80	1.11	.37	.21	.32	57.1	71.4	S3
10	11	15	-1.01	.62	1.09	.42	1.02	.23	.27	.32	71.4	71.4	S10
1	13	15	-1.96	.78	1.02	.21	1.17	.49	.18	.23	85.7	85.6	S1
MEAN	8.1	15.0	.00	.64	1.01	-.1	1.02	-.1			75.0	73.3	
P.SD	2.8	.0	1.07	.06	.30	.9	.45	.9			8.6	6.6	

Gambar 5. Tingkat kesulitan butir soal

Kelompok butir soal sukar yaitu S4 (1.59 logit), S5 dan S6 (1,13 logit). Kelompok butir soal sulit yaitu S7, S8 (0,37 logit). Kelompok butir soal mudah yaitu S2 dan S9 (-0,030 logit). Kelompok butir soal sangat mudah yaitu S3 S10 (-1, 01 logit) dan S1(-1,96 logit). Dengan demikian dapat diketahui tingkat kesulitan dari masing-masing soal secara nyata. Hal tersebut dikarenakan analisis tingkat kesulitan butir soal menggunakan Rasch menyajikan hasil analisis berdasarkan jawaban siswa terhadap soal, bukan berdasarkan distribusi persentase seperti pada analisis konvensional. Jika dikaitkan dengan indikator berpikir kritis pada setiap soal, maka butir soal yang cenderung sukar yaitu S4, S5 dengan indikator *reasoning*, dan S6 *analyzing*. Butir soal yang memiliki tingkat kesulitan sulit yaitu S7 dengan indikator *reasoning and claims*, dan S8 dengan indikator *judging*. Selanjutnya, butir soal mudah terdapat pada S2 dengan indikator *analyzing and claims*, dan S9 dengan indikator berpikir kritis *evaluating*. Kemudian, butir soal yang memiliki tingkat kesulitan sangat mudah yaitu S10 dengan indikator *claims*, dan S1 dengan indikator *analyzing*. Dengan demikian, meskipun dengan indikator yang sama, setiap soal memiliki tingkat kesulitan yang berbeda-beda.

Analisis lebih lanjut, mengenai kesesuaian butir soal dapat diketahui melalui nilai *outfit means-square*, *outfit z -standard*, dan *point measure correlation* (Bond & Fox, 2015; Boone, Yale, & Staver, 2014; Sumintono & Widhiarso, 2015). Kriteria yang digunakan untuk mengetahui kesesuaian butir soal yang tidak sesuai (*outliers* atau *misfit* ) adalah nilai *outfit* means-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-measure Corr*)  $0,4 < Pt-Measure Corr < 0,85$  (Boone et al., 2014; Sumintono & Widhiarso, 2015). Kesesuaian butir soal dari soal yang telah dikembangkan seperti pada gambar 6.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.		INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		Item
				MNSQ	ZSTD	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
4	4	15	1.59	.72	1.70	1.48	2.16	1.71	A	.17	.59	64.3	81.8	S4
6	5	15	1.13	.66	1.26	.82	1.28	.73	B	.40	.55	71.4	77.6	S6
3	11	15	-1.01	.62	1.19	.80	1.11	.37	C	.21	.32	57.1	71.4	S3
1	13	15	-1.96	.78	1.02	.21	1.17	.49	D	.18	.23	85.7	85.6	S1
10	11	15	-1.01	.62	1.09	.42	1.02	.23	E	.27	.32	71.4	71.4	S10
2	9	15	-.30	.58	.84	-.88	.76	-.60	e	.51	.41	78.6	65.3	S2
9	9	15	-.30	.58	.83	-.94	.76	-.61	d	.52	.41	78.6	65.3	S9
7	7	15	.37	.59	.78	-.99	.71	-.97	c	.63	.48	78.6	68.5	S7
8	7	15	.37	.59	.78	-.99	.71	-.97	b	.63	.48	78.6	68.5	S8
5	5	15	1.13	.66	.61	-1.19	.49	-1.31	a	.79	.55	85.7	77.6	S5
MEAN	8.1	15.0	.00	.64	1.01	-.1	1.02	-.1				75.0	73.3	
P.SD	2.8	.0	1.07	.06	.30	.9	.45	.9				8.6	6.6	

Gambar 6. Tingkat kesesuaian butir soal

Berdasarkan hasil analisis Rasch diperoleh tingkat kesesuaian butir soal yang merujuk pada kriteria dari nilai *outfit* MNSQ, *outfit* ZSTD dan *Pt-measure corr*. Seluruh soal memenuhi nilai *outfit* ZSTD, namun terdapat satu butir soal yang tidak memenuhi nilai *outfit* MNSQ yaitu S4 (1,70), dan terdapat empat soal yang tidak memenuhi nilai *Pt-measure corr* yaitu S4 (0,17), S3 (0,21), S1 (0,18) dan S10 (0,27). Meskipun demikian, seluruh butir soal dapat dipertahankan dan tidak perlu diubah atau diganti karena seluruh soal masih memenuhi nilai *outfit* ZSTD yang dibolehkan. Dengan demikian, seluruh butir soal dapat berfungsi normal dalam melakukan pengukuran atau dapat dikatakan fit.

Selanjutnya, dapat diketahui juga tingkat abilitas dari masing-masing siswa sehingga diperoleh informasi kemampuan siswa dalam mengerjakan soal tes tertulis berpikir kritis berbasis STEM. Seperti halnya tingkat kesulitan soal, data *logit* individu dapat menjelaskan banyak hal, karena skala yang dihasilkan mempunyai jarak yang sama (Sumintono & Widhiarso, 2015). Nilai *logit* yang tinggi menunjukkan tingkat kemampuan menyelesaikan yang tinggi. Tingkat abilitas 15 orang siswa yang telah mengerjakan soal tes tertulis yaitu seperti pada gambar 7.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-CORR.	AL-EXP.	EXACT MATCH		Person
					MNSQ	ZSTD	MNSQ	ZSTD			OBS%	EXP%	
13	10	10	3.88	1.83	MAXIMUM MEASURE								13P
8	9	10	2.58	1.09	.74	-.12	.35	-.15	.50	.26	90.0	89.8	08P
14	8	10	1.68	.84	.67	-.71	.44	-.53	.64	.35	80.0	79.7	14P
1	6	10	.52	.72	1.20	.74	1.09	.34	.30	.44	50.0	70.7	01P
6	6	10	.52	.72	1.44	1.41	2.20	2.21	-.05	.44	70.0	70.7	06L
9	6	10	.52	.72	.85	-.44	.89	-.10	.54	.44	90.0	70.7	09L
4	5	10	.02	.71	1.49	1.57	1.43	1.13	.07	.45	40.0	70.2	04P
12	5	10	.02	.71	.50	-2.01	.46	-1.72	.86	.45	100.0	70.2	12P
2	4	10	-.49	.72	1.26	.89	1.54	1.25	.19	.45	70.0	71.2	02L
5	4	10	-.49	.72	1.04	.23	1.07	.31	.41	.45	70.0	71.2	05L
7	4	10	-.49	.72	.72	-.91	.64	-.86	.69	.45	90.0	71.2	07L
10	4	10	-.49	.72	1.04	.23	1.07	.31	.41	.45	70.0	71.2	10P
15	4	10	-.49	.72	.56	-1.63	.48	-1.42	.82	.45	90.0	71.2	15L
3	3	10	-1.03	.77	1.49	1.31	2.16	1.74	-.06	.43	60.0	74.1	03L
11	3	10	-1.03	.77	.53	-1.48	.42	-1.14	.81	.43	80.0	74.1	11P
MEAN	5.4	10.0	.35	.83	.97	-.1	1.02	.1			75.0	73.3	
P.SD	2.1	.0	1.34	.28	.35	1.1	.60	1.1			16.4	5.2	

Gambar 7. Tingkat abilitas siswa (person measure)

Berdasarkan gambar 7 dapat diketahui siswa 13P memiliki abilitas paling tinggi yaitu dengan nilai *logit* 3,88 dengan total skor jawaban benar 10 soal. Kemudian, pada tingkat terendah yaitu siswa 3L dan 11P dengan nilai *logit* -1,03 hal ini menunjukkan kemampuannya yang jauh dari rata-rata person *logit* yaitu 0,35 *logit*. Dengan diketahui nilai dari masing-masing person *logit* maka dapat digunakan untuk memberikan perbandingan abilitas antar siswa. Nilai *logit* masing-masing siswa juga

dapat dikategorikan berdasarkan nilai standar deviasi dan rata-rata person *logit* yang diketahui. Abilitas siswa dapat dikategorikan menjadi tiga kategori, yaitu kataegori rendah, kategori sedang dan kategori tinggi. Kategori rendah ditentukan apabila siswa memiliki abilitas dengan nilai *logit* dibawah rata-rata person *logit* (0,35). Sehingga dapat diketahui yang berada pada kategori rendah yaitu 04P(0,02), 12P(0,02), 02L(-0,49), 05L(-0,49), 07L(-0,49),10P(-0,49),15L(-0,49), 03L(-1,03), dan 11P(-1,03),. Kategori sedang yaitu siswa dengan nilai *logit* diatas rata-rata person *logit* (0,35) namun masih dibawah nilai standar deviasi (1,34), siswa tersebut adalah 1P, 6L, 9L (0,52). Dan pada kategori tinggi yaitu siswa yang memiliki nilai *logit* diatas rata-rata person (0,35) dan diatas standar deviasi (1,34) yaitu 13P (3,88), 08P(2,58) dan 14P(1,68). Berdasarkan data tersebut, dapat diketahui abilitas siswa dalam berpikir kritis masih rendah, karena hanya 3 siswa yang memiliki kemampuan dalam kategori tinggi.

Selain itu, melalui pemodelan Rasch maka dapat dianalisis pula tingkat kesesuaian individu (person fit) dari masing-masing siswa. Tingkat kesesuaian individu ini dapat mendeteksi jika terdapat individu yang pola responnya tidak sesuai (Sumintono & Widhiarso, 2015). Pola respon dikatakan tidak sesuai jika jawaban yang diberikan oleh siswa tidak sesuai dengan abilitasnya. Tingkat kesesuaian individu menggunakan kriteria tingkat kesesuaian butir soal *outfit* means-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-measure Corr*)  $0,4 < Pt-Measure Corr < 0,85$  (Boone et al., 2014). Tingkat kesesuaian individu dapat dilihat pada gambar 8.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Person
					MNSQ	ZSTD	MNSQ	ZSTD					
6	6	10	.52	.72	1.44	1.41	2.20	2.21	A-.05	.44	70.0	70.7	06L
3	3	10	-1.03	.77	1.49	1.31	2.16	1.74	B-.06	.43	60.0	74.1	03L
2	4	10	-.49	.72	1.26	.89	1.54	1.25	C .19	.45	70.0	71.2	02L
4	5	10	.02	.71	1.49	1.57	1.43	1.13	D .07	.45	40.0	70.2	04P
1	6	10	.52	.72	1.20	.74	1.09	.34	E .30	.44	50.0	70.7	01P
5	4	10	-.49	.72	1.04	.23	1.07	.31	F .41	.45	70.0	71.2	05L
10	4	10	-.49	.72	1.04	.23	1.07	.31	G .41	.45	70.0	71.2	10P
9	6	10	.52	.72	.85	-.44	.89	-.10	g .54	.44	90.0	70.7	09L
8	9	10	2.58	1.09	.74	-.12	.35	-.15	f .50	.26	90.0	89.8	08P
7	4	10	-.49	.72	.72	-.91	.64	-.86	e .69	.45	90.0	71.2	07L
14	8	10	1.68	.84	.67	-.71	.44	-.53	d .64	.35	80.0	79.7	14P
15	4	10	-.49	.72	.56	-1.63	.48	-1.42	c .82	.45	90.0	71.2	15L
11	3	10	-1.03	.77	.53	-1.48	.42	-1.14	b .81	.43	80.0	74.1	11P
12	5	10	.02	.71	.50	-2.01	.46	-1.72	a .86	.45	100.0	70.2	12P
MEAN	5.4	10.0	.35	.83	.97	-.1	1.02	.1			75.0	73.3	
P.SD	2.1	.0	1.34	.28	.35	1.1	.60	1.1			16.4	5.2	

Gambar 8. Tingkat Kesesuaian Individu (Person Fit)

Berdasarkan gambar 8, dapat diketahui terdapat tujuh orang siswa yang tidak memenuhi nilai *outfit mean square* (MNSQ) yaitu 06L(2,20), 03L(2,16), 02L(1,54), 14P(0,44) 15L(0,48), 11P(0,42) dan 12P(0,46). Kemudian, terdapat satu orang siswa yang tidak memenuhi nilai *outfit z-standard* yaitu 06L(2,20). Selain itu, terdapat enam orang siswa yang tidak memenuhi nilai *point measure correlation* yaitu 06L(-0,5), 03L(-0,6), 02L(0,19), 04P(0,7), 01P(0,30), dan 12P(0,86). Dengan demikian, siswa-siswa tersebut dianggap tidak fit terutama siswa 06L yang tidak memenuhi ketiga kriteria tersebut dan memiliki nilai *point measure correlation* negatif yang menunjukkan adanya pola respon diluar kebiasaan. Untuk informasi lebih jauh dapat dilihat melalui skalogram Guttman seperti pada gambar 9.

Person	Item	
	1	
	1302978564	
	-----	
13	+1111111111	13P
8	+1111111110	08P
14	+1111111100	14P
1	+1111000101	01P
6	+0011111100	06L
9	+1101111000	09L
4	+1000111010	04P
12	+1111100000	12P
2	+1100010001	02L
5	+1010100010	05L
7	+1110001000	07L
10	+1010100010	10P
15	+1111000000	15L
3	+0101000001	03L
11	+1110000000	11P
	-----	
	1	
	1302978564	

Gambar 9. Guttman Scalogram

Berdasarkan skala guttman tersebut dapat diketahui bahwa siswa 13P, 08P,dan 14P memiliki respon yang sangat baik, karena mampu mengerjakan soal dari butir soal dari yang paling mudah ke soal yang paling sulit secara berurut. Berbeda dengan siswa 01P misalnya, siswa tersebut termasuk kategori siswa yang tidak cermat karena dapat mengerjakan soal yang sulit bahkan paling sulit namun tidak dapat menjawab soal yang tingkat kesulitannya sedang. Selanjutnya siswa 03L, siswa tersebut memberikan pola respon yang kurang baik, karena siswa tersebut mampu mengerjakan soal yang paing sulit namun tidak dapat mengerjakan soal paling mudah dengan benar.



Setiap siswa memiliki kemampuan yang berbeda-beda, begitupun butir soal yang dibuat dapat dikembangkan sehingga memiliki tingkat kesulitan yang beragam. Soal dapat dikembangkan sesuai dengan kebutuhan sebagaimana tuntutan keterampilan abad 21. Keterampilan tersebut dapat dilatih melalui mengerjakan soal-soal yang memiliki keterampilan berpikir tingkat tinggi. Pengembangan soal tes tertulis berbasis STEM dan memperhatikan keterampilan berpikir siswa menjadi solusi dari permasalahan dan kebutuhan pada saat ini. Hasil dari pengerjaan soal yang dianalisis memberikan informasi mengenai kualitas butir soal dan tingkat abilitas dari siswanya. Soal yang telah dibuat memiliki tingkat kesulitan yang beragam, sehingga dapat dikerjakan oleh berbagai siswa sesuai tingkat abilitasnya. Soal tersebut dapat menunjukkan kemampuan siswa dalam berpikir kritis berbasis STEM. Hasil analisis Rasch menunjukan tingkat kesulitan yang beragam namun memiliki tingkat kesesuaian yang kurang baik karena terdapat soal yang hanya memenuhi nilai ZSTD yang mengindikasikan terjadi miskonsepsi terhadap soal tersebut, sehingga perlu diperbaiki dalam proses pembelajarannya (Sumintono & Widhiarso, 2015). Hal tersebut yang menjadi kekurangan dan keterbatasan dalam penelitian ini. Hasil analisis yang diperoleh akan cenderung sama, jika soal diujikan kepada siswa dengan karakteristik cenderung pada level yang sama. Hasil dari analisis Rasch tersebut memberikan informasi untuk mengidentifikasi keterampilan berpikir kritis siswa berbasis STEM.

### ***Report***

Laporan disusun berdasarkan data yang diperoleh. Hasil akhir berupa 10 butir soal tes tertulis berbasis STEM dengan keterampilan berpikir kritis berdasarkan indikator "*Assessing 21<sup>st</sup> Century Skills for Teachers and Students*". Soal tes tertulis tersebut telah melalui analisis menggunakan pemodelan rasch sehingga dapat diketahui informasi mengenai kualitas butir soal berdasarkan tingkat kesulitan soal, tingkat kesesuaian soal, tingkat abilitas siswa dan tingkat kesesuaian siswa ketika mengerjakan soal tersebut.

## KESIMPULAN

Dapat disimpulkan bahwa hasil pada penelitian ini berupa 10 butir soal tes tertulis berbasis STEM dengan analisis pemodelan Rasch yang telah tervalidasi berdasarkan hasil focus group discussion bersama tim perangkat pembelajaran lainnya dan dinyatakan layak sebagai alat penilaian. Selain itu, hasil dari analisis menggunakan pemodelan rasch menunjukkan bahwa butir soal memiliki berbagai tingkat kesulitan yaitu sukar, sulit, mudah, dan sangat mudah untuk kemampuan siswa yang beragam (tinggi, sedang dan rendah). Oleh karena itu butir soal dinyatakan layak untuk digunakan menilai keterampilan berpikir kritis berbasis STEM.

## DAFTAR PUSTAKA

- Akhirani, Y. S. (2019). *Pengembangan Soal Tes Tertulis Pada Pembelajaran Berbasis Stem Untuk Mencapai Kemampuan Berpikir Kritis Dan Berpikir Kreatif Di Sekolah Dasar*. Universitas Pendidikan Indonesia.
- Ariyanti, P. L., Dantes, N., & Marhaeni, A. A. I. N. (2020). *Kelas I Berbasis Kecakapan Belajar Dan Berinovasi Abad 21*. 4(1), 147–158.
- Arseni, P. A. E., Lasmawan, W., & Suarni, K. (2020). ( RPP ) *Tema Tugasku Sehari-Hari Kelas Ii Sd Berbasis Kecakapan Belajar Dan Berinovasi Abad 21*. 4(1), 1–10.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model*. Taylor & Francis. Retrieved from <https://doi.org/10.4324/9781410614575>
- Boone, W. J., Yale, M. S., & Staver, J. . (2014). Rasch analysis in the human sciences. *Rasch Analysis in the Human Sciences*. Retrieved from <https://doi.org/10.1007/978-94-007-6857-4>
- Fitriani, D. (2018). Pengembangan Instrumen Tes Higher-Order Thinking Skill pada Pembelajaran Tematik Berbasis Outdoor Learning di Sekolah Dasar Kelas IV. *Indonesian Journal of Primary Education, Vol. 2*.
- Glynn, P., Shanahan, M., & Duggan, D. (2015). Focus groups. Retrieved from <https://www.slideshare.net/johnglynn940/focus-groups-presentation>
- Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). *Items Quality Analysis Using Rasch Model To Measure Elementary School Students ' Critical Thinking Skill On Stem Learning*. 9(1), 61–74. <https://doi.org/10.23887/jpi-undiksha.v9i1.20884>
- Hosnan. (2014). *Pendekatan Sainifik Dan Konstektual Dalam Pembelajaran Abad 21: Kunci Sukses Implementasi Kurikulum 2013*. bogor: Penerbit Ghalia.

- Lidinillah, D., Mulyana, E., Karlimah, & Hamdu, G. (2019). Integration of STEM learning into the elementary curriculum in Indonesia: An analysis and exploration Integration of STEM learning into the elementary curriculum in Indonesia: An analysis and exploration. *Journal of Physics*. <https://doi.org/10.1088/1742-6596/1318/1/012053>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2015). *TIMSS 2015 International Results in Mathematics*.
- OECD. (2019). PISA 2018 insights and interpretations. *OECD Publishing*, 64. Retrieved from [https://www.oecd.org/pisa/PISA 2018 Insights and Interpretations FINAL PDF.pdf](https://www.oecd.org/pisa/PISA_2018_Insights_and_Interpretations_FINAL_PDF.pdf)
- Saputra. (2016). *Pengembangan Mutu Pendidikan Menuju Era Global: Penguatan Mutu Pembelajaran dengan Penerapan HOTS (High Order Thinking Skills)*. Bandung: SMILE's Publishing.
- Sumintono, B. (2018). Pengukuran Model Rasch sebagai Alat Penilaian untuk Belajar. *Kemajuan Dalam Ilmu Sosial, Pendidikan Dan Penelitian Humaniora*, 173 (Icei. Retrieved from <https://doi.org/10.2991/icei-17.2018.11>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch Pada Assesment Pendidikan*. Cimahi: Penerbit Trim Komunikata.
- Trilling & Fadel. (2009). 21st Century Skills. *Learning For Life in Our Times*. Jossey Bass : USA.
- Zulfikar, D. (2019). *Pengembangan Instrumen Penilaian Kinerja Pada Pembelajaran Berbasis Stem Untuk Mencapai Kemampuan 4C Di Sekolah Dasar*. Universitas Pendidikan Indonesia Kampus Tasikmalaya.