

Pengembangan Instrumen *Higher Order Thinking Skill* Mata Pelajaran Matematika Dengan *Rasch Model*

Heni Hadiyanti¹, Purwo Susongko², Munadi³

(1) Program Studi Magister Pedagogi, Pascasarjana, Universitas Pancasakti Tegal

(2) Program Studi Magister Pedagogi, Pascasarjana, Universitas Pancasakti Tegal

(3) Program Studi Magister Pedagogi, Pascasarjana, Universitas Pancasakti Tegal

✉ Corresponding author
henialiyahediyanti@gmail.com

Abstrak

Higher Order Thinking Skill (HOTS) peserta didik di SMK Ma'arif NU Talang masih rendah. Penelitian ini bertujuan untuk: (1) Mengidentifikasi kebutuhan tes-tes Matematika bagi Guru dan Peserta Didik; (2) Membangun konstruksi tes Mata Pelajaran Matematika berbasis HOTS; (3) Menguji validitas aspek isi tes Mata Pelajaran Matematika berbasis HOTS; (4) Menguji kualitas aspek psikometri butir-butir tes Mata Pelajaran Matematika berbasis HOTS; dan (5) Menguji validitas konstruk tes Mata Pelajaran Matematika berbasis HOTS. Untuk mencapai tujuan tersebut peneliti melaksanakan penelitian menggunakan pendekatan *Analysis, Design, Development*. Hasil penelitian menunjukkan bahwa: (1) terdapat kebutuhan pengembangan tes Matematika bagi Guru dan Peserta Didik; (2) Konstruksi tes telah mengacu pada indikator soal HOTS; (3) Instrumen tes telah memenuhi validitas dari aspek isi; (4) instrumen tes telah memenuhi validitas aspek psikometrik; dan (5) hasil analisis validitas konstruk menunjukkan bahwa terdapat 2 butir soal tidak terpakai dan 18 butir soal layak dimanfaatkan untuk mengukur HOTS.

Kata Kunci: *Higher Order Thinking Skill* (Hots), *Rasch Model*

Abstract

The Higher Order Thinking Skill (HOTS) of students at SMK Ma'arif NU Talang is still low. This research aims to: (1) Identify the need for Mathematics tests for Teachers and Students; (2) Building a HOTS-based Mathematics subject test construction; (3) Testing the validity of the content aspects of the HOTS-based Mathematics Subject test; (4) Testing the quality of the psychometric aspects of the HOTS-based Mathematics Subject test items; and (5) Testing the construct validity of the HOTS-based Mathematics Subject test. To achieve this goal, researchers carried out research using the *Analysis, Design, Development* approach. The research results show that: (1) there is a need to develop Mathematics tests for Teachers and Students; (2) The test construction refers to the HOTS question indicators; (3) The test instrument has met the validity of the content aspect; (4) the test instrument has met the validity of psychometric aspects; and (5) the results of the construct validity analysis show that there are 2 unused questions and 18 questions that are suitable for measuring HOTS.

Keyword: *Higher Order Thinking Skill* (Hots), *Rasch Model*

PENDAHULUAN

Proses evaluasi hasil pembelajaran tidak hanya bermanfaat untuk peserta didik, tetapi juga berguna bagi guru atau pendidik sebagai fasilitator pembelajaran. evaluasi pembelajaran dapat dimanfaatkan guru dalam merancang dan mengimplementasikan program pembelajaran selanjutnya agar sesuai dengan tujuan yang ingin dicapai. Kemampuan guru dalam mengelola pembelajaran (kompetensi pedagogik) memberi gambaran tentang keterampilan khas guru jika dibandingkan dengan profesi lain. Keterampilan tersebut menentukan tingkat keberhasilan proses dan hasil yang akan dicapai peserta didik sebagai citra kesuksesan proses pembelajaran (Hatta, 2018: 79).

Evaluasi hasil pembelajaran tidak hanya bermanfaat untuk peserta didik, tetapi juga berguna bagi guru atau pendidik sebagai fasilitator pembelajaran. Salah satu manfaat hasil evaluasi pembelajaran adalah dapat dimanfaatkan oleh guru dalam merancang dan mengimplementasikan program pembelajaran selanjutnya agar sesuai dengan tujuan yang ingin dicapai. Kemampuan guru dalam mengelola pembelajaran atau biasa disebut sebagai kompetensi pedagogik memberi gambaran mengenai keterampilan khas guru dibanding

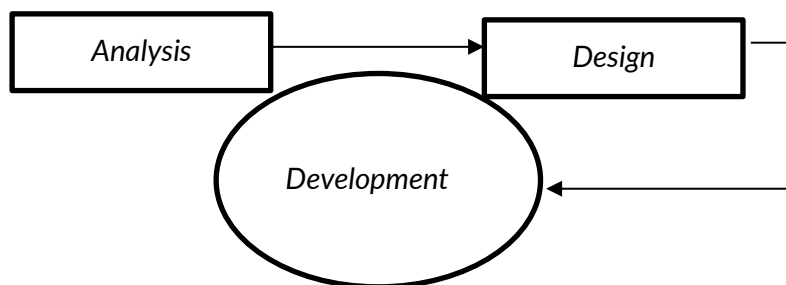
profesi lain. Keterampilan tersebut akan menentukan tingkat keberhasilan proses dan hasil yang akan dicapai peserta didik sebagai citra kesuksesan proses pembelajaran yang dilaksanakan.

Salah satu alternatif yang dapat dilakukan untuk memperbaiki kelemahan konsep evaluasi pembelajaran dapat dilakukan dengan mengembangkan sistem evaluasi dengan pendekatan lain. Rasch Model merupakan model analisis yang dikembangkan oleh Dr. Georg Rasch dan dapat dimanfaatkan untuk mengungkap tingkat kemampuan butir soal dalam mengukur kemampuan peserta didik didasarkan pada tingkat kesulitan soal yang dijadikan sebagai instrumen tes. Untuk memperoleh gambaran mengenai hal tersebut data diubah dalam bentuk 'odds ratio' melalui algoritma yang disesuaikan menjadi unit logit sehingga dapat memberi gambaran mengenai respon peserta didik dalam memberikan respon terhadap item soal. Hasil dari Rasch Model memberi informasi holistik mengenai instrumen soal serta dapat mengungkapkan unsur-unsur lain berdasarkan tes yang dilakukan.

Aryadoust et al., (2021) dalam penelitiannya menunjukkan manfaat Rasch Model secara bertahap meningkat dan sering dimanfaatkan untuk meningkatkan keterampilan menulis, berbicara, dan terpadu. Penelitian ini bertujuan untuk mengembangkan instrumen Mata Pelajaran Matematika agar terjadi peningkatan HOTS peserta didik di SMK Ma'arif NU Talang. Sehingga hipotesis yang akan dibuktikan dalam penelitian ini adalah Rasch Model efektif untuk mengembangkan instrumen tes Mata Pelajaran Matematika agar terjadi peningkatan HOTS peserta didik di SMK Ma'arif NU Talang.

METODE PENELITIAN

Penelitian ini merupakan Penelitian Pengembangan atau Research and Development (R & D) dan dilaksanakan di SMK Ma'arif NU Talang Kabupaten Tegal. Objek penelitian ini adalah instrument penilaian berbasis instrumen Higher Order Thinking Skill dengan Rasch Model pada Mata Pelajaran Matematika melalui tahap Research, Development, dan Diffusion. Sampel dalam penelitian ini adalah Siswa Kelas XI TKR SMK Ma'arif NU Talang. Teknik yang digunakan dalam penentuan sampel pada penelitian ini menggunakan Purposive Sampling. Pengumpulan Data pada penelitian ini menggunakan 2 teknik yaitu dokumentasi dan tes.



Produk penelitian ini adalah instrument penilaian berbasis instrumen Higher Order Thinking Skill dengan model Rasch Model pada Mata Pelajaran Matematika. Dalam penyusunan instrumen dilakukan uji validitas untuk menjamin kesesuaian instrumen yang dikembangkan dengan tujuan yang ingin dicapai. Dalam penyusunan instrumen Higher Order Thinking Skill atau kemampuan berfikir tingkat tinggi menggunakan indikator HOTS berdasarkan taksonomi Bloom revisi. Indikator dari masing-masing aspek yang dikembangkan sesuai Taksonomi Bloom Revisi.

HASIL DAN PEMBAHASAN

Analisis Kebutuhan

Model pengembangan Tes Mata Pelajaran Matematika berbasis HOTS dengan pendekatan Rasch Model pada penelitian ini terdiri dari lima tahap, yaitu Perencanaan Instrumen Tes, Uji Coba Instrumen, Penetapan Validitas, Penetapan Reliabilitas, dan Interpretasi Skor. Setelah instrumen Tes disusun dan diujicoba, selanjutnya dilakukan analisis untuk mengetahui validitas Tes berdasarkan skor ujicoba. Uji Validitas yang dilakukan dalam penelitian ini meliputi Uji Validitas Isi, Uji Validitas Psikometri, dan Uji Validitas Empiris. Validitas Isi dan Psikometri diperoleh berdasarkan penilaian *Expert* (ahli), sedangkan Validitas Empiris diperoleh berdasarkan analisis mengenai kualitas, karakteristik butir soal, dan kesimpulan mengenai kelayakan instrumen Tes yang telah dikembangkan. Produk akhir yang dikembangkan dalam

penelitian ini adalah Tes Mata Pelajaran Matematika berbasis HOTS yang terdiri dari 20 butir soal pilihan ganda.

Tahap perencanaan instrumen meliputi penentuan materi, aspek/ dimensi yang dinilai, dan indikator yang digunakan dalam pengembangan Tes Matematika berbasis HOTS. Materi yang menjadi bahan penyusunan instrumen pada penelitian ini merupakan materi ajar yang telah diberikan kepada Peserta Didik Kelas XI pada semua program kejuruan di SMK Ma'arif NU Talang. Sedangkan dimensi atau aspek yang dinilai melalui instrumen penelitian meliputi tiga aspek dalam HOTS yaitu menganalisis, menilai, dan mencipta. Indikator yang digunakan dalam penyusunan instrumen soal merupakan penyesuaian antara materi dan aspek yang akan diukur dalam penelitian ini.

Peneliti mengujicobakan instrumen Tes kepada peserta didik Kelas XI semua program keahlian yang berperan sebagai responden dalam penelitian ini. Hasil dari ujicoba instrumen Tes tersebut digunakan untuk mengidentifikasi kebutuhan pengembangan Tes berdasarkan Indikator yang telah ditetapkan. Sebelum membuat Instrumen Tes peneliti terlebih dahulu membuat Angket analisis kebutuhan instrument HOTS dan divalidasi oleh para ahli. Hasil Validasi Isi Angket Kebutuhan Instrument HOTS tersebut adalah sebagai berikut:

Tabel 1 Rubrik Validasi Isi Angket Kebutuhan Instrument HOTS

No	Rubrik	Penilai 1		Penilai 2	
		Ya	Tidak	Ya	Tidak
1	Kejelasan petunjuk pengisian angket	✓		✓	
2	Cakupan isi pernyataan / pertanyaan mendeskripsikan tujuan angket yang ditetapkan	✓		✓	
3	Penggunaan bahasa komunikatif (mudah dimengerti), pernyataan / pertanyaan spesifik dan tidak menimbulkan penafsiran ganda	✓		✓	

Hasil Analisis kebutuhan per Indikator berdasarkan Tes ujicoba yang dilakukan pada 50 guru menggunakan angket *Google Form*.

Tabel 2 Analisis kebutuhan per Indikator

No	Indikator	Prosentase
1	<i>Analyzing</i> (menganalisis)	47,00%
2	<i>Evaluating</i> (menilai)	42,00%
3	<i>Creating</i> (mencipta)	35,00%
4	HOTS	49,00%

Sumber : Diolah dari data primer dengan R Program

Hasil analisis kebutuhan per Indikator ditemukan bahwa perolehan skor responden pada semua indikator HOTS berada di bawah 50% atau rendah, sehingga dibutuhkan pengembangan instrumen Tes untuk meningkatkan keterampilan berpikir tingkat tinggi (HOTS) peserta didik.

Analisis Materi

Terdapat tiga dimensi HOTS, yaitu Menganalisis (C4), Mengevaluasi (C5), dan Mencipta (C6). Menganalisis (C4) merupakan keterampilan menguraikan gagasan ke bagian yang lebih kecil dan rinci untuk mendapatkan makna secara mendalam. Secara praktis HOTS meliputi proses transfer, berpikir kritis, dan menyelesaikan masalah. HOTS mendorong terjadinya proses belajar bermakna sehingga peserta didik mampu menerapkan hasil belajar pada situasi dan kondisi yang beragam. HOTS juga mendorong peserta didik untuk dapat berpikir secara logis dan reflektif, sehingga memiliki kemampuan mengambil keputusan dan menyelesaikan masalah riil secara mandiri. Untuk mempermudah penyusunan soal HOTS dibutuhkan Kata Kerja Operasional (KKO).

Butir soal HOTS dalam bentuk pilihan ganda bertujuan untuk menguji pemahaman peserta didik mengenai masalah secara komprehensif. Materi soal HOTS tersebut berbentuk pilhan ganda kompleks yang di dalamnya terdapat stimulan berupa situasi kontekstual berupa bacaan. Agar tujuan pengukuran keterampilan berpikir tingkat tinggi tercapai, maka pilihan jawaban disusun secara acak dan tidak mengikuti pola tertentu.

Design

Desain pengembangan yang dilakukan dalam penulisan soal HOTS pada penelitian ini meliputi empat tahap: (1) menentukan materi yang akan dinilai; (2) menyusun kisi-kisi; (3) merumuskan indikator soal; dan (4) menulis soal sesuai dengan kaidah penulisan soal. Instrumen HOTS mencakup 5 materi dan 3 dimensi yang terdiri berfikir kreatif, berfikir kritis dan pemecahan masalah, serta mencakup 3 level kognitif C4 (menganalisis), C5 (mengevaluasi) dan C6 (mencipta).

Development Aspek Konten

Validasi Aspek Konten dalam penelitian ini dilakukan oleh para ahli di bidangnya dan diperoleh hasil sebagai berikut:

Tabel 3 Rubrik Aspek Konten Instrumen Higher Order Thinking Skill (HOTS)

No	Rubrik	Penilai 1		Penilai 2	
		Ya	Tidak	Ya	Tidak
1	Narasi sesuai dengan domain capaian pembelajaran pada mata pelajaran Matematika	√		√	
2	Setiap butir soal mengandung minimal satu dimensi Higher Order Thinking Skill (HOTS)	√		√	
3	Kunci jawaban benar	√		√	

Sumber : Data Primer

Hasil Validasi Isi pada menunjukkan instrumen Tes Matematika yang dikembangkan untuk meningkatkan keterampilan berpikir tingkat tinggi (HOTS) dalam penelitian ini dinyatakan layak ditinjau dari Aspek Isi.

Validasi Aspek Psikometrik

Validasi aspek psikometrik bertujuan memastikan butir-butir tes memenuhi kaidah psikometrik dalam penyusunan butir. Dari hasil penilaian pakar di bidang psikometrik disimpulkan bahwa instrument HOTS pada mata pelajaran Matematika yang telah dibuat layak untuk diuji coba.

Validitas Konstruk Aspek Isi

Validitas Isi mengukur kemampuan instrumen dalam mengukur hal yang ingin diukur. Hasil Validitas Isi adalah penilaian tentang kelayakan isi instrumen. Semakin detail penjelasan mengenai instrumen alat ukur mengindikasikan instrumen tersebut semakin tepat dengan tujuan pengukuran. Uji coba instrument dilaksanakan pada tanggal 10 Oktober 2023 di SMK Ma'arif NU Talang. Hasil analisis dengan model berisi hasil analisis kecocokan butir terhadap model (*item fit*).

Tabel 4 Hasil Analisis Item Fit Instrumen HOTS Siswa Kelas XI SMK Ma'arif NU Talang.

No. Butir	Chisq	Df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t	Discrim
V1	163.373	215	0.996	0.756	0.897	-1.360	-0.712	0.255
V2	225.356	215	0.300	1.043	0.989	0.439	-0.110	-0.575
V3	211.917	215	0.547	0.981	0.956	-0.063	-0.332	-0.035
V4	170.711	215	0.988	0.790	0.883	-1.476	-1.076	0.390
V5	175.089	215	0.979	0.811	0.896	-1.108	-0.792	0.325
V6	228.776	215	0.247	1.059	0.996	0.433	0.003	-0.277
V7	204.627	215	0.683	0.947	0.969	-0.662	-0.489	-0.135
V8	194.497	215	0.839	0.900	0.967	-0.613	-0.249	-0.067
V9	179.148	215	0.964	0.829	0.929	-0.958	-0.503	0.176
V10	309.793	215	0.000	1.434	0.946	1.334	-0.148	-0.295
V11	175.632	215	0.977	0.813	0.867	-1.484	-1.359	-0.016
V12	166.757	215	0.994	0.772	0.900	-0.692	-0.353	0.103
V13	220.876	215	0.377	1.023	0.906	0.174	-0.485	0.141
V14	188.656	215	0.902	0.873	0.886	-1.376	-1.604	0.034
V15	155.670	215	0.999	0.721	0.880	-0.724	-0.346	0.297
V16	143.540	215	1.000	0.665	0.863	-0.990	-0.449	0.520
V17	231.852	215	0.205	1.073	0.971	0.367	-0.088	-0.215

No. Butir	Chisq	Df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t	Discrim
V18	160.851	215	0.998	0.745	0.889	-0.981	-0.510	0.456
V19	190.792	215	0.881	0.883	0.914	-0.286	-0.291	0.099
V20	173.666	215	0.982	0.804	0.914	-0.937	-0.513	0.328

Sumber : Data Primer

Item fit pada dasarnya menjelaskan apakah suatu butir berfungsi melakukan pengukuran secara normal atau tidak. Secara kuantitatif butir tes yang dinyatakan fit atau dapat berfungsi dengan baik jika nilai *Outfit MSQ* antara 0.5 hingga 1.5, sedangkan nilai *Outfit t* antara -2 hingga 2 serta peluang H_0 (kecocokan model) lebih besar dari 0.05 ($p > 0.05$).

Seluruh butir secara umum dapat diterima sebagai butir yang baik kecuali butir nomor 10. Butir nomor 10 memiliki $p\text{-value} < 0.01$ dapat disimpulkan bahwa taraf signifikan 0.01 butir nomor 10 tidak bisa diterima oleh model. Item Tes disebut memiliki tingkat kesukaran yang rendah apabila $b < 2,0$ logit; tingkat kesukaran sedang jika $-2,0 \text{ logit} \leq b \leq 2,0 \text{ logit}$; dan tingkat kesukaran tinggi jika $b > 2,0$ logit. Tingkat kesukaran soal berdasarkan analisis data pada penelitian ini adalah -1,559 sampai 3,121.

Tabel 5 Nilai Tingkat Kesukaran Butir-butir Instrumen HOTS

No. butir	Tingkat Kesukaran	No. butir	Tingkat Kesukaran
1	-0.369	11	3.121
2	2.828	12	-1.257
3	-0.136	13	-0.762
4	0.000	14	2.674
5	-0.248	15	-1.559
6	-0.033	16	-1.449
7	2.441	17	-0.884
8	-0.066	18	-0.949
9	-0.287	19	-1.257
10	-1.257	20	-0.550

Sumber : Diolah dari data primer dengan R Program

Tingkat kesukaran paling rendah (-1,559) merupakan item Tes nomor 15, tingkat kesukaran paling tinggi adalah item Tes pada nomor 11 dengan skor tingkat kesukaran adalah 2.828. Tingkat kesukaran sebesar 2.828. mempunyai arti bahwa peserta diharapkan dapat mengerjakan butir dengan benar jika mempunyai kemampuan minimal -2.828.

Berdasarkan kriteria yang telah disebutkandan hasil analisis data diperoleh informasi bahwa terdapat 15 butir item Tes yang termasuk kateogri rendah, 1 item Tes termasuk kategori sedang, dan 4 item Tes yang termasuk kategori sulit pada pengembangan butir soal Matematika yang disusun dalam penelitian ini. Parameter tingkat kesulitan butir digambarkan oleh suatu titik pada skala kemampuan dimana peluang menjawab benar sebesar 0,5. Semakin besar nilai parameter tingkat kesulitan , maka semakin besar kemampuan yang dibutuhkan responden untuk mendapatkan peluang menjawab butir soal dengan benar sebanyak 0,5.

Tingkat kesukaran butir bergerak dari -3.169 hingga 2.938. Tes yang efektif , memiliki tingkat kesukaran butir antara -2.00 hingga 2.00 (Wright, & Stone, 1979; Hambleton, et al, 1991; Wu & adam, 2007). Namun demikian tes yang dibangun untuk mengukur kompetensi seperti halnya Instrumen HOTS pada mata pelajaran matematika untuk siswa SMK sebaiknya dapat mengukur kemampuan semua peserta tes sehingga sebaran tingkat kesukarannya lebih luas dari tes yang dibangun dalam paradigma tes seleksi

atau tes yang menggunakan acuan norma. Bila diasumsikan seperti halnya yang dikembangkan oleh item response theory/ distribusi normal, maka tingkat kesukaran butir untuk pengukuran yang bersifat kompetensi dapat dimulai dari - 3.00 hingga 3.00, karena pada interval tersebut dapat mengukur sekitar 95 % peserta tes.

Hasil analisis semua butir tes pada interval interval -3.00 hingga 3.00 sehingga efektif sebagai tes kompetensi. Analisis tersebut menunjukkan bahwa butir instrument HOTS pada mata pelajaran matematika siswa SMK efektif digunakan untuk kemampuan peserta tes antara -3.00 hingga 3.00, sehingga hipotesis (ha) diterima.

Validitas Konstruk Aspek Substantif

Secara kuantitatif respons peserta tes yang dinyatakan fit atau tidak mengalami penyimpangan adalah bila nilai Outfit MSQ antara 0.5 hingga 1.5 sedangkan nilai outfit t antara -2 hingga 2,0 serta peluang penerimaan H_0 (kecocokan model) lebih besar dari 0.01 ($p > 0.01$).

Hasil uji person fit dari 216 jawaban peserta tes HOTS pada mata pelajaran matematika di SMK Ma'arif NU Talang. ada 18 peserta tes yang mengalami respon yang menyimpang dari model. Hal ini terlihat ke 18 peserta tes tersebut tidak memenuhi sebanyak dua (p value dan outfit MSQ) dari ketiga kriteria person fit, Bahkan ada 7 peserta tes (P1, P52, P56, P124, P206, P211, dan P214) tidak memenuhi seluruh kriteria person fit. Daftar peserta tes tersebut paparkan dalam tabel berikut.

Tabel 6 Peserta Tes Yang Memiliki Respons Menyimpang (*aberrant response*)

Peserta	Chisq	Df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P1	90.566	19	0.000	4.528	2.866	2.87	2.77
P15	49.137	19	0.000	2.457	1.822	2.00	1.60
P34	73.138	19	0.000	3.657	1.855	1.94	1.65
P52	45.886	19	0.001	2.294	1.885	2.17	1.90
P56	56.121	19	0.000	2.806	2.318	2.32	2.27
P77	54.065	19	0.000	2.703	1.370	1.30	0.87
P120	46.313	19	0.000	2.316	1.501	1.54	1.06
P121	44.455	19	0.001	2.223	1.774	1.77	1.53
P124	42.810	19	0.001	2.141	1.797	2.20	2.05
P144	47.948	19	0.000	2.397	1.352	1.34	0.84
P155	52.576	19	0.000	2.629	2.262	1.77	2.10
P163	42.244	19	0.002	2.112	1.754	1.66	1.50
P171	47.161	19	0.000	2.358	1.236	1.32	0.63
P187	60.820	19	0.000	3.041	1.959	1.67	1.79
P206	54.266	19	0.000	2.713	1.832	2.24	1.61
P207	40.161	19	0.003	2.008	2.196	1.55	2.11
P211	73.451	19	0.000	3.673	2.817	2.43	2.72
P214	112.381	19	0.000	5.619	1.963	2.10	1.79

Sumber : Diolah dari data primer dengan R Program

Terdapat 91.67% respons peserta tes yang wajar dan sesuai dengan model, sedangkan respon peserta tes yang mengalami penyimpangan sebesar 8.33%. Besarnya prosentase peserta tes yang memiliki respons wajar mengindikasikan bahwa tes telah memenuhi validitas substantif.

Validasi Konstruk Aspek Struktural

Berdasarkan Uji unidimensi Instrumen butir-butir Instrumen Tes pada penelitian ini diperoleh hasil sebagai berikut:

Tabel 7 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.489
Approx. Chi-Square	10.479

Bartlett's Test of Sphericity	Df	10
	Sig.	.399

Sumber : Diolah dari data primer dengan R Program

Angka KMO *Measure of sampling Adequacy* (MSA) adalah 0.489. Hal ini menunjukkan kecukupan dari sampel. Angka KMO dan Bartlett's test (yang tampak pada nilai *chi-square*) sebesar 10.479 dengan nilai signifikansi 0.399. Nilai tersebut menunjukkan bahwa terdapat korelasi antar variabel sehingga pemanfaatan instrumen layak untuk digunakan pada proses analisis berikutnya.

Untuk mengetahui variabel yang dapat diproses lebih lanjut dan variabel yang harus dikeluarkan agar memenuhi syarat komunalitas, nilai *extraction* harus lebih dari 0,50. Terdapat 3 faktor yang terbentuk dari 5 variabel atau butir soal yang dimasukkan. Nilai *eigenvalue* menggambarkan kepentingan relative dari setiap faktor dalam varians dari 5 variabel atau butir soal yang di analisis. Besarnya Varians yang mampu dijelaskan oleh faktor baru yang terbentuk adalah 65.567% sedangkan sisanya 34.443% dijelaskan oleh faktor lain yang tidak diteliti. Nilai *eigenvalue* pada 3 faktor tersebut lebih dari 1.

Untuk melakukan uji invariansi pengukuran menggunakan uji *Anderson LR test*. Uji ini digunakan untuk mengetahui konsistensi estimasi parameter pemodelan Rasch. Kondisi ideal pada pemodelan Rasch terjadi jika estimasi parameter tingkat kesukaran item konsisten (invarian) meski diperoleh dari sampel yang terdiri dari subkelompok populasi mana selama menerapkan pemodelan Rasch, dalam hal ini menggunakan PCM. Hasil analisis *Anderson LR test* diperoleh nilai *p value* sebesar 0.011 artinya menerima H_0 sehingga dapat disimpulkan bahwa estimasi parameter bersifat invarian. Hal tersebut bermakna bahwa respon terhadap item tes tidak akan mengalami perubahan meski diterapkan pada situasi yang berbeda.

Validasi Konstruk Aspek Eksternal

Data yang diuji validitas kontruk terdiri dari nilai perolehan hasil uji instrumen HOTS dan nilai Penilaian Akhir Semester (PAS). Syarat yang harus dipenuhi agar Validitas Konstruk Aspek Eksternal dapat diterima adalah jika nilai Cronbach's Alpha lebih dari 0,60 atau 60%, dan hasil analisis data menunjukan hasil nilai Cronbach's Alpha sebesar 0,685 sehingga dinyatakan memenuhi validasi konstruk aspek eksternal.

Pentingnya *Rasch Model* sebagai alat analisis yang mengukur instrumen Tes dan peserta didik dalam mengembangkan keterampilan berpikir tingkat tinggi (HOTS) didasarkan pada dua alasan. Alasan pertama adalah instrumen Tes yang baik akan menjadi percuma jika dalam pengerjaannya dilakukan secara tidak jujur melalui perilaku menyimpang peserta didik. Sedangkan alasan yang kedua adalah instrumen yang buruk tidak akan memberi makna terhadap hasil yang dicapai, meski peserta didik telah mengerjakan soal dengan jujur dan baik. Oleh sebab penggunaan uji validitas *Rasch Model* bertujuan agar intrumen Tes hasil pengembangan dalam penelitian ini dapat memberi gambaran ketepatan item soal sekaligus diperoleh informasi mengenai karakteristik peserta didik dalam mengerjakan instrumen Tes yang digunakan. Hal tersebut dapat diketahui berdasarkan item fit dan person fit hasil analisis validitas yang diterapkan.

Ditinjau dari Validitas Isi yang berfungsi untuk mengukur relevansi instrumen tes dalam mencapai tujuan yang ditetapkan, hasil analisis statistik menunjukan bahwa instrumen Tes pengembangan dapat mengukur hal yang ingin diukur. Hal tersebut terlihat dari *statistik outfit* yang berada pada rentang 0.756 sampai 1,434 atau $0,5 \leq MSQ \leq 1,5$. Menurut *rule of thumb* yang disampaikan oleh Linacre (2002), implikasi kecocokan model terhadap pengukuran dapat dilihat berdasarkan MSQ. Jika $MSQ > 2,0$ bermakna merusak sistem pengukuran; $1,5 < MSQ \leq 2,0$ mengandung arti tidak memiliki makna terhadap pengukuran; $0,5 \leq MSQ \leq 1,5$ berarti bermanfaat bagi pengukuran; dan $MNSQ < 0,5$ berarti tidak bermanfaat terhadap pengukuran meski tidak tes tersebut tidak merusak sistem pengukuran. Semua item instrumen Tes pengembangan dalam penelitian ini bermanfaat bagi pengukuran dan terbukti fit dengan model Rasch, serta tidak ada *item misfit* atau item Tes yang berkontribusi besar terhadap keandalan dalam pengukuran kemampuan.

Hasil analisis validasi konstruk aspek isi pada penelitian ini, dari 216 peserta tes tampak bahwa seluruh butir secara umum dapat diterima sebagai butir yang baik kecuali butir nomor 10. Butir nomor 10 memiliki *p-value* < 0.01 dapat disimpulkan bahwa taraf signifikan 0.01 butir nomor 10 tidak bisa diterima oleh model. Nilai tingkat kesukaran tidak normal ada pada butir nomor 11 sebesar 3.121 sedangkan tingkat kesukaran tertinggi pada butir nomor 2.828. Berdasarkan hasil analisis Tingkat Kesukaran Soal pada instrumen tes diketahui bahwa tingkat kesukaran soal berada pada kisaran -1,559 sampai 3,121, karena

berada di antara -2,0 sampai 2,0 . Tingkat kesukaran paling rendah sebesar -1,559 merupakan item Tes nomor 15, sedangkan tingkat kesukaran paling tinggi sebesar 3,121 ada pada item Tes nomor 11. Terlihat pada butir soal nomor 10, 12, 15, 16 dan 19 memiliki nilai tingkat kesukaran yang rendah. Secara statistik dan ilustrasi Item Map Instrumen Tes pengembangan yang disusun memiliki item-item yang dapat digunakan untuk meningkatkan keterampilan berpikir tingkat tinggi (HOTS) karena berada tingkat Sedang dan Tinggi.

Hasil analisis Validitas Konstrak Aspek Substantif terdapat 18 peserta atau 8.33% dinyatakan menyimpang. Hal itu juga menunjukkan bahwa 91,67% peserta Tes telah memenuhi Validasi Aspek substantif berdasarkan Model Rasch. Hasil analisis tersebut juga mengandung arti bahwa 91,67% peserta Tes mengerjakan soal secara jujur dan tidak melakukan tindakan menyimpang yang dapat mengganggu kesimpulan, sehingga hasilnya dapat dipertanggung jawabkan.

Validitas aspek struktural berdasarkan Uji unidimensi butir-butir Instrumen Tes pada penelitian ini memiliki *nilai eigen value* yang kedua dan seterusnya lebih kecil dari *nilai eigen value* yang pertama. Kondisi tersebut menegaskan bahwa tes hanya mengandung satu dimensi, sehingga instrumen Tes pada penelitian ini bersifat unidimensi dalam mengukur keterampilan berpikir tingkat tinggi (HOTS).

Hasil Validitas Konstrak Aspek Eksternal terlihat bahwa nilai person separation reliability sebesar -0.0934. Dengan demikian nilai separasi person untuk tes tersebut adalah 0.458. Dari nilai separasi person tersebut dapat diketahui bahwa klasifikasi tes yang di dapat termasuk soal berfikir tingkat tinggi (HOTS).

Dari hasil penelitian diketahui bahwa ada dua butir soal yang tidak layak digunakan sebagai instrument pengukuran kemampuan berfikir tingkat tinggi (HOTS) yaitu butir yang tidak cocok dengan model (butir soal nomor 10) dan butir soal yang melebihi range tingkat kesukaran yaitu butir soal nomor 11. Sementara butir yang lain dengan analisis validitas yang mencakup isi, psikometri dan konstrak memenuhi persyaratan sebagai butir yang baik.

SIMPULAN

Berdasarkan hasil analisis menggunakan Rasch Model dapat disimpulkan bahwa terdapat kebutuhan pengembangan tes Matematika bagi Guru dan Peserta Didik di SMK Ma'arif NU Talang untuk meningkatkan keterampilan berpikir tingkat tinggi (HOTS). Hasil analisis validitas kontrak tes Mata Pelajaran Matematika berbasis HOTS dalam penelitian ini diperoleh hasil bahwa 2 butir soal tidak terpakai dan 18 butir soal layak untuk digunakan, sehingga dapat dimanfaatkan untuk mengukur keterampilan berpikir tingkat tinggi (HOTS).

DAFTAR PUSTAKA

- Abduh, M. (2019). *Panduan Penulisan Soal HOTS-Higher Order Thinking Skills*. Jakarta: Pusat Penilaian Pendidikan.
- Agus Budiman., J. (2014). Pengembangan Instrumen Asesmen Higher Order Thinking Skill (Hots) Pada Mata Pelajaran Matematika Smp Kelas Viii Semester 1. *Jurnal Riset Pendidikan Matematika, Volume 1 - Nomor 2, November 2014*, 139-151.
- Anisa. (2013). Perbandingan Penskoran Dikotomi dan Politomi dalam Rasch Model untuk Pengembangan Bank Soal Mata Kuliah Matematika Dasar. *Jurnal Matematika, Statistika, & KOMputasi* Vo. 9. No. 2. Januari 2013, 95-113.
- Artha Mahindra Diputera. (2018). *Analisis IRT Menggunakan Wingen 3: Rasch Model & Aplikasi*. Ponorogo: Uwais Inspirasi Indonesia.
- Cindy Priscilla., D. Y. (2021). Implementasi Pilar-Pilar Pendidikan UNESCO. *Asatiza: Jurnal Pendidikan n Vol 2 No 1 (2021)*, 64-76.
- Didana, I. W. (2017). *Modul Penyusunan Soal Higher Orer Thingking Skill (HOTS)*. Jakarta: Direktorat Pembinaan Sekolah Menengah Atas Ditjen Pendidikan Dasar dan Menengah, Kementerian Pendidikan dan Kebudayaan.
- DPR RI. (2016). *Permendikbud Nomor 23 Tahun 2016 Tentang Standar Penilaian Pendidikan*. Jakarta: DPR RI.
- Fradia Mayang Intan, E. K. (2020). Kemampuan Siswa dalam Mengerjakan Soal HOTS (Higher Order Thinking Skills) pada Pembelajaran Matematika di Kelas V Sekolah Dasar. *Jurnal Pendidikan Dasar Indonesia Volum 5 Nomor 1 bulan Maret 2020* , 6-10.
- Haryanto. (2020). *Evaluasi Pembelajaran (Konsep dan Manajemen)*. Yogyakarta: UNY Press.
- Huriaty, D. (2015). Metode Kalibrasi Dan Desain Tes Berdasarkan Teori Respons Butir (IRT). *Math Didactic: Jurnal Pendidikan Matematika Vol. 1, No. 3, September - Desember 2015*, 191-199.

- Komarudin., S. (2017). *Evaluasi Pembelajaran*. Yogyakarta: Rizqita Publishing & Printing.
- Kemendikbud Dirjen Diksar dan Menengah . (2017). *Pengembangan butir soal Higher-Order Thinking Skills Higher-Order Thinking Skills(HOTS)*. Jakarta: Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Dasar dan Menengah Direktorat Pembinaan Sekolah Menengah Atas.
- Maryani,Ika (2022). *Pengembangan Pembelajaran IPA Model Mishe (Metacognition in Science for Higher Education) Untuk Meningkatkan Higher-Order Thinking Skills Mahasiswa*.
- Melly Elvia, S. H. (2015). Karakter BUTir Soal Ujian Semester dan Kemampuan Siswa SMA di Kabupaten Muaro Jambi. *Jurnal Evaluasi Pendidikan* Vol. 4. NO.1, Maret 2015, 58-68.
- PISA Governing Board Indonesia. (2019). *Pendidikan Di Indonesia Belajar dari hasil PISA 2018*. Jakarta : Pusat pengembangan pendidikan bidang penelitian dan pendidikan Kmendikbud.
- Retnawati, H. (2014). *TEORI RESPONS BUTIR DAN PENERAPANNYA Untuk Peneliti, Praktisi Pengukuran dan PengujianJ Mahasiswa Pascasarjana*. Yogyakarta: Nuha Madika.
- Rohim, D. C. (2019). Strategi Penyusunan Soal Berbasis HOTS pada Pembelajaran Matematika SD. *BRILIANT: Jurnal Riset dan Konseptual Volume 4 Nomor 4, November 2019*, 436-446.
- Sainuddin, S. (2018). Analisis Karakter Butir Tes Matematika Pada Tes Buatan MGMP Matematika KOTA Makassar Berdasarkan Teori MODeren (Rasch Model). *PROXIMA: Jurnal Penelitian Matematika dan Pendidikan Matematika*, 1-164.
- Sugiharto. (2016). Perbedaan Ketidakwaian Skor Berdasarkan Rasch Model Ditinjau Dari Model Penskoran Jawaban Siswa Smp Se Kota Palangka Raya. *Anterior Jurnal, Volume 16 Nomor 1, Desember 2016* , 28-39.
- Supriadi, G. (2020). *Pengembangan Instrumen Penilaian Berbasis Higher Order Thinking Skill (HOTS)*. Sleman: Aswaja Pressindo.
- Susongko,P.(2016).*Aplikasi Model Rasch dalam Pengukuran Pendidikan Berbasis Program R*. Tegal:Badan Penerbitan Universitas Pancasakti Tegal (BPUPS)
- Susongko,P.(2017).*Penilaian Hasil Belajar*. Tegal:Badan Penerbitan Universitas Pancasakti Tegal (BPUPS)
- Tim Pusat Penilaian Pendidikan. (2019). *Panduan Penulisan Soal HOTS-Higher Order Thinking Skills*. Jakarta: Pusat Penilaian Pendidikan.
- Winaryati. E., M. M. (2021). *Cercular Model of RD&D (Model RD&D Pendidikan dan Sosial)*. Bantul: Penerbit KBM Indonesia
- Wiwik Setiawati, d. (2018). *Buku Pegangan Pembelajaran Berorientasi pada Keterampilan Berpikir Tingkat Tinggi*. Jakarta: Direktorat Jenderal Guru dan Tenaga Kependidikan Kemendikbud.
- Wahid, A. H. (2018). Integrasi higher order thinking skill (hots) dengan model creative problem solving. *Modeling: Jurnal Program Studi PGMI Volume 5, Nomor 1, Maret 2018*, 82-98.