

Pengembangan Asesmen *Higher Order Thinking Skill (HOTS)* Pada Mata Pelajaran IPAS Berbasis Model *Rasch* Tingkat Sekolah Dasar

Sudiryo^{1✉}, Sitti Hartinah², Purwo Susongko³
(1,2,3) Pedagogi, Universitas Pancasakti Tegal

✉ Corresponding author
[sudiryo2020@gmail.com]

Abstrak

Asesmen yang dilaksanakan guru sering kali tidak memperhatikan aspek sejauh mana soal yang digunakan mampu mengukur keterampilan berpikir siswa tingkat tinggi (*High Order Thinking Skills*). Masih banyak guru tingkat SD dalam penyusunan soal tes hanya bersifat hafalan dan berfokus pada aspek *kognitif* C1-C3 (mengukur betul atau salah). Soal yang ada lebih mengedepankan pada pengertian suatu istilah dalam bidang ilmu, belum menggali lebih dalam penerapan konsep dalam kehidupan sehari-hari. Selain itu, guru tingkat SD belum melakukan analisis butir soal menggunakan model tertentu (*Rasch*). Penelitian ini bertujuan untuk 1) menganalisis kebutuhan guru dan siswa terkait pengembangan asesmen *HOTS* mata pelajaran IPAS kelas IV, 2) mengetahui desain pengembangan asesmen *HOTS* mata pelajaran IPAS kelas IV, 3) menguji validitas butir tes pada aspek isi dan aspek psikometri, serta 4) menguji validitas konstruk butir tes berbasis model *Rasch*. Penelitian dilakukan menggunakan desain penelitian dan pengembangan dalam membangun butir tes *HOTS* mata pelajaran IPAS tingkat Sekolah Dasar. Butir tes disusun, selanjutnya diuji validitasnya. Pengujian validitas meliputi uji validitas aspek isi, psikometri, serta konstruk. Uji validitas isi dan psikometri yang dilakukan oleh ahli menyatakan bahwa butir tes yang disusun layak untuk diuji cobakan. Selanjutnya, butir tes diujicobakan kepada 153 peserta siswa dari 5 sekolah sampel di Gugus Ki Hajar Dewantara Kecamatan Talang Kabupaten Tegal. Hasil uji coba tes diuji validitas konstruk menggunakan pemodelan *Rasch* dengan *r programming* 4.3.3 yang meliputi uji validitas konstruk aspek isi, substansi, struktural, dan eksternal. Hasil validitas konstruk dengan pemodelan *Rasch* sebagai berikut: (1) Tingkat kesukaran butir berada pada range -5 hingga 5, artinya butir-butir tes tersebut cocok untuk semua kemampuan peserta tes, (2) Pada taraf kepercayaan 95%, ada 19 butir yang cocok dengan pemodelan, (3) Pada taraf kepercayaan 95% sebanyak 93,46% respons siswa cocok dengan pemodelan, (4) Pada taraf kepercayaan 100%, tidak ada butir yang mengandung DIF. Dengan demikian, asesmen *HOTS* mata pelajaran IPAS yang disusun dapat digunakan untuk mengukur kemampuan *HOTS* siswa secara valid berdasarkan pemodelan *Rasch*.

Kata kunci: Asesmen, *HOTS*, IPAS, *ADDIE*, Model *Rasch*

Abstract

Assessments carried out by teachers often do not pay attention to the aspect of the extent to which the questions used are able to measure students' high-order thinking skills. There are still many elementary school teachers in the preparation of test questions only by rote and focus on the cognitive aspects of C1-C3 (measuring right or wrong). The existing problems prioritize the meaning of a term in the field of science, not more in the application of concepts in daily life. In addition, elementary school teachers have not analyzed question items using a certain model (*Rasch*). This study aims to 1) analyze the needs of teachers and students related to the

development of HOTS assessments for science subjects in class IV, 2) find out the design of HOTS assessment development for science subjects in class IV, 3) test the validity of test items in the content and psychometric aspects, and 4) test the validity of test item constructs based on the Rasch model. The research was carried out using research and development design in building HOTS test items for science and science subjects at the elementary school level. Test items are prepared, then tested for validity. Validity testing includes testing the validity of content aspects, psychometrics, and constructs. The validity test of the content and psychometrics carried out by experts stated that the test items prepared were worthy of testing. Furthermore, the test items were tested on 153 student participants from 5 sample schools in the Ki Hajar Dewantara Cluster, Talang District, Tegal Regency. The results of the test test were tested for construct validity using Rasch modeling with r programming 4.3.3 which includes testing the validity of the construct in terms of content, substance, structural, and external aspects. The results of the validity of the construct with Rasch modeling are as follows: (1) The difficulty level of the items is in the range of -5 to 5, meaning that the test items are suitable for all abilities of the test participants, (2) At the confidence level of 95%, there are 19 items that match the modeling, (3) At the confidence level of 95%, as many as 93.46% of the students' responses are suitable for modeling, (4) At the confidence level of 100%, none of the grains contain DIF. Thus, the HOTS assessment of science subjects that is prepared can be used to validly measure students' HOTS abilities based on Rasch modeling.

Keywords: *Assessment, HOTS, IPAS, ADDIE, Rasch Model*

PENDAHULUAN

Menurut Lie, A (2020) menjelaskan keterampilan berpikir tingkat tinggi adalah suatu kemampuan dalam menggunakan dan mengolah proses berpikir di atas fakta. Seseorang yang mempunyai keterampilan berpikir tingkat tinggi tidak hanya mengetahui suatu fakta tertentu, tetapi juga menggunakan pengetahuan yang diperoleh untuk mengembangkan pengetahuan itu sendiri. Keterampilan berpikir tingkat tinggi bisa diajarkan dan dilatihkan. Para pendidik harus selalu mengembangkan kemampuan dan melatih diri dalam mengolah keterampilan tersebut agar bisa mendampingi siswa berpikir lebih baik dan mengembangkan kebiasaan tersebut. Proses pembelajaran yang memberi suasana tantangan bagi siswa untuk menemukan fakta baru, dapat melatih kemampuan berpikir tingkat tinggi anak.

Tujuan utama dari *High Order Thinking Skills* adalah bagaimana meningkatkan keterampilan berpikir siswa pada tingkat yang lebih tinggi, terutama pada hal yang berhubungan dengan kecakapan untuk berpikir secara kritis dan kreatif dalam berbagai jenis informasi yang diterima, sehingga mampu diaplikasikan dalam memecahkan suatu permasalahan dengan menggunakan pengetahuan yang dimilikinya serta membuat keputusan dalam situasi-situasi yang kompleks. Salah satu upaya untuk meningkatkan *Higher Order Thinking Skills* pada siswa adalah dengan melakukan asesmen atau penilaian. Asesmen merupakan proses yang dilakukan dalam kegiatan yang sistematis dalam rangka mengumpulkan informasi tentang sesuatu, misalnya tentang perkembangan siswa atau kemajuan belajar yang dicapai.

Instrumen soal tes yang akan di ujikan ke siswa belum dilakukan uji validitas, reliabilitas dan obyektivitas sehingga kualitas instrumen yang dihasilkan belum mampu untuk mengukur kemampuan siswa secara *holistik*. Salah satu cara untuk melakukan uji validitas, reliabilitas dan obyektivitas instrumen soal adalah dengan pemodelan *rasch*. Dari uraian latar belakang di atas, peneliti mengemukakan beberapa tujuan penelitian adalah:

1. Untuk mengetahui kebutuhan guru dan siswa terkait pengembangan asesmen *HOTS* mata pelajaran IPAS kelas IV di tingkat Sekolah Dasar.
2. Untuk mengetahui desain pengembangan asesmen *HOTS* mata pelajaran IPAS kelas IV di tingkat Sekolah Dasar.
3. Untuk mengetahui hasil validitas aspek isi, validitas aspek psikometri pada pengembangan asesmen *HOTS* mata pelajaran IPAS kelas IV di tingkat Sekolah Dasar.
4. Untuk mengetahui hasil validitas konstruk dengan pemodelan *Rasch* pada pengembangan asesmen *HOTS* mata pelajaran IPAS kelas IV di tingkat Sekolah Dasar.

METODE PENELITIAN

Penelitian yang dilakukan menggunakan jenis *Riset and Development* (penelitian dan pengembangan) dengan model ADDIE yang dibatasi sampai tahap pengembangan dengan pemodelan *rasch*. Dalam upaya memperoleh gambaran terkait kebutuhan pengembangan asesmen HOTS mata pelajaran IPAS di Dabin 4 Gugus Ki Hajar Dewantara, peneliti melakukan wawancara dengan 3 siswa setiap SD sampel, Guru kelas IV Sd Sampel dan Kepala Sekolah SD Sampel. Uji validasi yang dilakukan peneliti melalui uji validasi isi, validasi psikometri dan validasi konstruk meliputi:

1. Isi (*content*)
Validitas isi menunjukkan bahwa semua butir dalam tes atau tugas yang melibatkan proses kognitif untuk menjawab betul sesuai dan mewakili dari bidang konstruk yang diukur. Aspek isi dari validitas konstruk berkaitan dengan tiga hal yaitu keseuaian isi, keterwakilan dan kualitas teknis.
2. Subtansi (*substantive*)
Berkaitan dengan subtansi dari aspek isi, dengan menemukan secara empirik untuk menjamin bahwa pengambilan tes secara actual benar-benar melibatkan kemampuan bidang yang diukur dalam menjawab butir soal. Sebagai contoh dalam tes pilihan ganda maka pengambilan tes yang memilih jawaban salah (*distractor*) benar-benar mempunyai kemampuan yang rendah.
3. Struktur (*structural*)
Struktur berkaitan dengan penskoran, skor pada tes multidimensi harus dilaporkan terpisah sesuai dengan dimensi masing-masing.
4. Generalisasi (*generalizability*)
Generalisasi untuk mengkaji sejauh mana skor yang diperoleh benar-benar menunjukkan kemampuan yang sebenarnya dari pengambil tes.
5. Eksternal (*external*)
Mengkaji sejauh mana skor yang diperoleh dari tes berkorelasi dengan tes lain yang sesuai.
6. Konsekuensi (*consequential*)
Berkaitan dengan pemaknaan dari skor yang diperoleh.

Tabel. Kriteria Tes Yang Valid Dilihat dari Berbagai Aspek Validitas dan Kriterianya Dengan Penerapan Model Rasch

Aspek Validasi Aspek	Indikator	Kriteria
Isi	Uji kecocokan item (<i>itemfit</i>)	$P > 0.01$ $0,5 \text{MNSQ} < 1,5$ $-2,0 < \text{ZSTD} < 2,0$
	<i>Person-item Map</i>	Semua tingkat kesukaran item berada pada domain kemampuan testee
	<i>Person/item Map</i>	Kemampuan testee sama atau mendekati tingkat kesukaran item
	Fungsi Informasi tes	Fungsi informasi tes mempunyai nilai maksimal pada domain kemampuan testee
Substantif	<i>Person fit statistic</i>	$P > 0.01$ $0,5 < \text{MNSQ} < 1,5$ $-2,0 < \text{ZSTD} < 2,0$
	<i>Collapsed Deviance/Casewise-Lemeshow</i>	$P < 0,01$
	<i>accuracy, sensitivity, specificity</i> dan	mendekati 1,0
Struktural	Uji unidimensi	ada satu faktor utama yang digambarkan lewat Scree Plot hasil analisis faktor
	Uji Invariansi	$P < 0,01$

Aspek Validasi Aspek	Indikator	Kriteria
	(LRtest)	
Eksternal	nilai separation	mendekati 1,0
	Person strata	
Konsekuensial	DIF	tidak terdapat DIF yang signifikan

Populasi adalah wilayah generalisasi yang terdiri atas: obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya (Sugiyono, 2017). Populasi dalam penelitian ini adalah siswa tingkat SD kelas IV Daerah Binaan (Dabin) 4 gugus Ki Hajar Dewantara Kecamatan Talang Kabupaten Tegal . Daftar populasi secara rinci di sajikan dalam tabel berikut:

Tabel. Jumlah Siswa Kelas IV Dabin 4 Gugus Ki Hajar Dewantara

No	Nama SD	Laki-laki	Perempuan	Jumlah
1	SD Negeri Benge 01	16	10	26
2	SD Negeri Gembong 01	18	15	33
3	SD Negeri Gembong 02	10	14	24
4	SD Negeri Pasangan 01	13	9	22
5	SD Negeri Pasangan 02	17	7	24
6	SD Negeri Pegirikan 01	19	12	31
7	SD Negeri Pegirikan 02	9	7	16
8	SD Negeri Pegirikan 03	22	14	36
9	SD Negeri Pekiringan 01	15	19	34
10	SD Negeri Pekiringan 02	13	12	25
	Jumlah	152	119	271

Sumber : Dapodik 2024

Sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi. Untuk itu sampel yang diambil dari populasi harus betul- betul representatif atau mewakili (Sugiyono,2017). Peneliti menggunakan teknik pemilihan sampel dengan *Purposive Sampling* dengan pertimbangan sebagai berikut :

- Keterwakilan SD setiap desa di wilayah yang ada di gugus Ki Hajar Dewantara Kecamatan Talang.
- Jumlah siswa paling banyak dari setiap SD yang desa wilayah gugus Ki Hajar Dewantara untuk memenuhi kestabilan uji konstruk pemodelan *Rasch*.

Tabel. Jumlah Sampel penelitian

No	Nama SD	Laki-laki	Perempuan	Jumlah
1	SD Negeri Benge 01	16	10	26
2	SD Negeri Gembong 01	18	15	33
3	SD Negeri Pasangan 02	17	7	24
4	SD Negeri Pegirikan 03	22	14	36
5	SD Negeri Pekiringan 01	15	19	34
	Jumlah	88	65	153

HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan sesuai dengan tahapan ADDIE adalah sebagai berikut :

Analysis (Analisis)

Analisis merupakan upaya untuk memperoleh gambaran faktual dilapangan terkait proses pelaksanaan asesmen yang dilakukan guru di SD sampel yang telah ditentukan peniliti. Proses yang dilakukan peneliti adalah dengan datang langsung ke SD kemudian menggali informasi melalui siswa, guru kelas IV dan Kepala sekolah. semua siswa menjawab guru melakukan asesmen dengan mengerjakan soal yang ada di buku paket atau soal yang ditulis guru di papan tulis.

Wawancara peneliti dengan siswa meliputi 8 point pertanyaan dengan hasil pertanyaan nomor satu menjelaskan bahwa semua guru kelas melakukan asesmen di akhir pembelajaran. Jawaban pertanyaan nomor dua, semua siswa menjawab soal yang diberikan guru berbentuk campuran (pilihan ganda dan isian). Jawaban pertanyaan nomor tiga sebanyak 10 siswa (67%) menjawab tidak tahu apakah soal sudah sesuai tujuan pembelajaran dan 5 siswa (33%) menjawab tahu atau soal sudah sesuai dengan tujuan pembelajaran. Jawaban siswa atas pertanyaan tingkat kesulitan soal (nomor empat) yang diberikan guru, sebanyak 5 siswa (33%) menjawab kebanyakan mudah, 5 siswa (33%) menjawab ada yang mudah dan ada yang susah dan 5 siswa (33%) menjawab soal yang diberi susah. Untuk pertanyaan nomor lima, sebanyak 2 siswa (13%) menjawab bisa mengukur pemahaman terhadap materi yang di pelajari dan 13 siswa (97%) menjawab tidak tahu atau tidak paham.

Sebanyak 2 siswa (13%) menjawab suka soal berbentuk isian tetapi 13 siswa (97%) menjawab menyukai soal berbentuk pilihan ganda.

Hasil wawancara dengan guru kelas IV Sebagian besar mereka belum membuat kisi-kisi soal saat Menyusun instrument asesmen karena mereka menggunakan soal yang ada di buku pake atau buku bahan ajar. Guru kelas IV juga belum melakukan analisis tingkat kesulitan soal, namun mereka menggunakan hasil asesmen formatif sebagai tolak ukur keberhasilan pemahaman siswa terhadap materi yang dipelajari. Dari hasil supervisi yang dilakukan oleh kepala sekolah, para guru mendapat masukan untuk selalu memperhatikan tingkat keterampilan berpikir tingkat tinggi (HOTS) dalam penyusunan instrument soal.

Hasil wawancara dengan Kepala sekolah diperoleh gambaran perlunya pengembangan asesmen yang mampu mengukur tingkat keterampilan berpikir tingkat tinggi (HOTS) mata Pelajaran IPAS. Bercermin dari kebutuhan tersebut, Kepala Sekolah menghendaki guru Kelas IV mampu membuat instrument soal sesuai CP dan tingkat kognif HOTS.

Hasil uji aspek isi yang dilakukan oleh dua orang pakar atau ahli (Prof. Dr. Sitti Hartinah. DS, MM. dan Prof. Dr. Purwo Susongko, M.Pd) menunjukkan bahwa instrument soal yang dibuat peneliti layak digunakan karena telah memenuhi aspek validasi isi sebagai berikut :

Tabel 4.6 Hasil Validasi Isi Instrumen Asesmen HOTS IPAS

No	Aspek yang dinilai	Hasil Penilaian			
		Ahli 1		Ahli 2	
		Ya	Tidak	Ya	Tidak
1	Narasi sesuai dengan domain capaian pada mata pelajaran IPAS	✓		✓	
2	Narasi berbasis fenomena atau kejadian yang sering dialami dalam kehidupan sehari-hari	✓		✓	
3	Setiap item soal mengandung tingkat kemampuan atau keterampilan berfikir (kognitif) tingkat tinggi (C4-C6)	✓		✓	
4	Kunci jawaban benar	✓		✓	

Hasil uji aspek psikometri yang dilakukan oleh Rosyidin, S.Pd.SD.,M.Pd. selaku ahli bidang asesmen karena beliau termasuk tim validasi naskah soal PAS dan PAT di KWK Dikbud Kecamatan Talang dan ahli dalam bidang IPA Fahmi Fatkhomi, M.Pd. menunjukkan instrument asesmen HOTS mata pelajaran IPAS yang disusun peneliti telah memenuhi unsur-unsur dalam psikometri meliputi materi, kontruksi dan bahasa . Hasil uji aspek psikometri tersaji dalam tabel berikut:

Tabel 4.7 Hasil Validasi Psikometri instrumen HOTS IPAS (1-10)

No	Aspek yang dinilai	Hasil Penilaian	
		Ahli 1	Ahli 2
Materi			
1	Soal sesuai indikator (Capaian Pembelajaran)	Memenuhi	Memenuhi
2	Pilihan jawaban homogen dan logis dari segi materi	Memenuhi	Memenuhi
3	Setiap item soal mempunyai jawaban yang benar dan paling benar	Memenuhi	Memenuhi
Kontruksi			
4	Pokok soal dirumuskan secara jelas dan tegas	Memenuhi	Memenuhi
5	Rumusan pokok soal dan jawaban merupakan pernyataan yang diperlukan	Memenuhi	Memenuhi
6	Pokok soal jangan memberikan arah ke jawaban yang paling benar	Memenuhi	Memenuhi
7	Pokok soal tidak mengandung pernyataan negatif bersifat ganda	Memenuhi	Memenuhi
8	Panjang rumusan soal relatif sama	Sebagian memenuhi	Memenuhi
9	Pilihan jawaban tidak mengandung pernyataan : "Semua pilihan jawaban di atas benar" "Semua pilihan jawaban di bawah benar"	Memenuhi	Memenuhi
10	Pilihan jawaban yang berbentuk angka atau waktu disusun berdasarkan urutan besar kecilnya nilai angka tersebut.	Sebagian memenuhi	Memenuhi

Tabel 4.8 Hasil Validasi Psikometri instrumen HOTS IPAS (11-16)

No	Aspek yang dinilai	Hasil Penilaian	
		Ahli 1	Ahli 2
11	Soal berbentuk tabel, grafik dan diagram tertulis jelas dan berfungsi	Memenuhi	Memenuhi
12	Butir soal tidak tergantung pada jawaban soal sebelumnya	Memenuhi	Memenuhi
<i>Bahasa</i>			
13	Setiap soal menggunakan kaidah bahasa yang benar	Memenuhi	Memenuhi
14	Tidak menggunakan bahasa yang berlaku sempit, jika soal akan digunakan untuk daerah lain atau nasional.	Memenuhi	Memenuhi
15	Setiap soal harus menggunakan bahasa yang komunikatif	Memenuhi	Memenuhi
16	Pilihan jawaban jangan mengulang kata atau frase yang bukan merupakan satu kesatuan pengertian	Memenuhi	Memenuhi

Uji validasi konstruk melalui pemodelan *Rasch* meliputi :

Tabel 4.10 Hasil analisa Item Fit Instrumen Asesmen HOTS IPAS (1-10)

No Butir	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
1	158.725	152	0.338	1.037	1.038	0.492	0.640
2	105.130	152	0.999	0.687	0.878	-0.839	-0.372
3	158.883	152	0.335	1.038	1.037	0.586	0.690
4	150.506	152	0.519	0.984	1.001	-0.247	0.032
5	136.227	152	0.816	0.890	0.936	-1.006	-0.792
6	124.702	152	0.949	0.815	0.870	-0.755	-0.729
7	170.941	152	0.140	1.117	1.109	1.292	1.576
8	146.207	152	0.617	0.956	0.971	-0.680	-0.553
9	153.393	152	0.453	1.003	1.001	0.063	0.038
10	137.510	152	0.794	0.899	0.940	-1.588	-1.151

Tabel 4.11 Hasil analisa Item Fit Instrumen Asesmen HOTS IPAS (11-20)

No Butir	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
11	154.109	152	0.437	1.007	1.032	0.112	0.503
12	121.281	152	0.968	0.793	0.863	-1.904	-1.672
13	155.784	152	0.400	1.018	0.983	0.241	-0.247
14	129.338	152	0.909	0.845	0.962	-0.786	-0.230
15	160.817	152	0.297	1.051	1.021	0.729	0.394
16	142.969	152	0.688	0.934	0.953	-1.050	-0.908
17	150.218	152	0.526	0.982	0.980	-0.238	-0.349
18	155.989	152	0.396	1.020	0.988	0.225	-0.137
19	92.493	152	1.000	0.605	0.864	-1.429	-0.562
20	186.921	152	0.028	1.222	1.076	1.695	0.836

Item fit pada dasarnya menjelaskan apakah suatu butir berfungsi melakukan pengukuran secara normal atau tidak. Syarat 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 2,0 serta peluang penerimaan H_0 (kecocokan model) lebih besar dari 0,01 ($p > 0,01$). *Outfit* adalah *outlier-sensitive fit*, yaitu suatu ukuran kesensitifan pola respons terhadap item dengan tingkat kesulitan tertentu dari para responden (siswa) atau sebaliknya. Ketidakkcocokan respon dengan model bisa disebabkan oleh banyak faktor misalnya adanya kecerobohan, miskonsepsi atau keberhasilan menebak. Nilai *Outfit MSQ* dihitung dari nilai *chi square* dibagi dengan derajat kebebasan (*df*). Dari Tabel 4.5 tampak bahwa seluruh butir secara umum dapat diterima sebagai butir yang baik.

Parameter tingkat kesukaran butir soal apabila nilai dari P sebagai berikut: $P \leq -2$ (butir soal dikategorikan sangat mudah), $-2 \leq P \leq -1$ (butir soal dikategorikan mudah), $-1 \leq P \leq 1$ (butir soal dikategorikan sedang), $1 \leq P \leq 2$ (butir soal dikategorikan sukar), dan $P \geq 2$ (butir soal dikategorikan sangat sukar). Besarnya nilai tingkat kesukaran pada setiap butir soal dapat dilihat berikut:

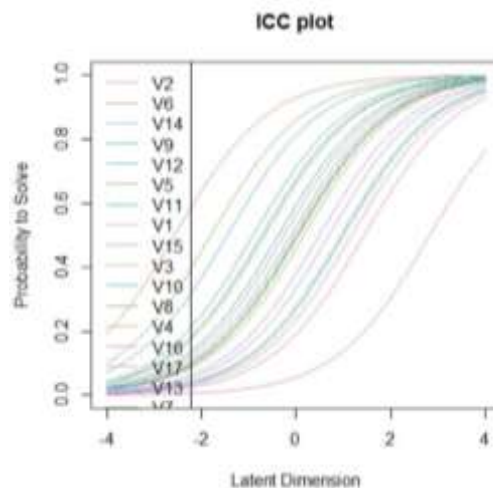
Tabel 4.12 Nilai Tingkat Kesukaran Butir-butir Instrumen Asesmen HOTS IPAS

Butir soal	Nilai	Keterangan
1	-0.9498	Sedang
2	-5.3586	Sangat mudah
3	-0.5558	Sedang
4	-0.2233	Sedang
5	-1.6175	Mudah
6	-3.7671	Sangat mudah
7	1.2788	Sukar
8	-0.4104	Sedang
9	-2.0566	Sangat mudah
10	-0.4548	Sedang
11	-1.1531	Mudah
12	-1.7153	Mudah
13	1.1308	Sukar
14	-3.0349	Sangat mudah
15	-0.7005	Sedang
16	0.3243	Sedang
17	0.6694	Sedang
18	1.5222	Sukar
19	5.1191	Sangat sukar
20	2.0194	Sangat sukar

Tabel di atas menunjukkan perbedaan pada setiap butir soalnya dengan tingkat kesukaran tertentu. Berdasarkan Tabel 4.6 terdapat 4 butir soal tergolong sangat mudah, 3 butir soal

tergolong mudah, 8 butir soal tergolong sedang, 3 butir soal tergolong sukar, dan 2 butir soal tergolong sangat sukar. Persentasi kategori pada tingkat kesukaran butir soal masing-masing yaitu: 20% soal dikategorikan sangat mudah, 15% soal dikategorikan mudah, 40% soal dikategorikan sedang, 15% soal dikategorikan sukar, dan 10% soal dikategorikan sangat sukar. Nilai tingkat kesukaran berkisar antara -5.3586 sampai dengan 5.1191.

Untuk lebih jelasnya Gambar 4.1 menjelaskan kurva karakteristik dari semua butir-butir soal.



Gambar 4.1 Kurva Karakteristik Butir-Butir Instrumen Asesmen HOTS IPAS

Gambar kurva di atas menjelaskan bahwa semakin tinggi kemampuan responden, maka semakin naik pula peluang menjawab benarnya. Butir nomor 19 adalah butir yang paling sukar disusul kemudian oleh butir nomor 20, 18, 7, dan seterusnya, dan paling mudah adalah butir nomor 2.

Respon yang menyimpang ini dapat ditimbulkan oleh adanya ketidakcermatan, adanya kegiatan *cheating* (menyontek) atau bahkan adanya miskonsepsi. Uji respons seseorang mengalami penyimpangan atau tidak disebut *person fit*. Kriteria penerimaan respons peserta tes dianggap mengalami penyimpangan atau tidak sama dengan kriteria *item fit*. Secara kuantitatif respon peserta tes yang dinyatakan fit atau tidak mengalami penyimpangan apabila nilai *Outfit MSQ* antara 0.5 hingga 1.5 sedangkan nilai *outfit t* antara -2.0 hingga 2.0 serta peluang penerimaan H_0 (kecocokan model) lebih besar dari 0.05 ($p > 0.05$). Ketiga kriteria tersebut harus dipenuhi untuk menyimpulkan terjadinya *abberant response*, jika salah satu tidak terpenuhi maka dianggap tidak terjadi *abberant response*.

Tabel 4.11 Hasil Uji person fit instrument asesmen HOTS IPAS

Peserta	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P1	30.341	19	0.048	1.517	1.356	1.45	1.73
P2	16.284	19	0.638	0.814	0.891	-0.49	-0.54
P3	20.643	19	0.357	1.032	1.026	0.21	0.20
P4	10.461	19	0.941	0.523	0.806	-0.37	-0.32
P5	11.347	19	0.912	0.567	0.717	-0.72	-0.96
P6	15.108	19	0.716	0.755	0.843	-0.70	-0.81
P7	19.163	19	0.446	0.958	1.074	-0.02	0.45
P8	18.962	19	0.459	0.948	1.083	0.01	0.43
P9	18.726	19	0.475	0.936	0.989	-0.09	0.00
P10	43.579	19	0.001	2.179	1.640	2.32	2.50
P11	22.260	19	0.272	1.113	1.267	0.39	0.81
P12	19.340	19	0.435	0.967	1.112	0.09	0.50

Peserta	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P13	8.857	19	0.976	0.443	0.569	-1.35	-1.95
P14	19.319	19	0.437	0.966	1.010	0.02	0.12
P15	12.465	19	0.865	0.623	0.715	-1.25	-1.65
P16	28.975	19	0.066	1.449	1.264	1.11	1.17
P17	13.948	19	0.787	0.697	0.787	-0.83	-1.08
P18	14.751	19	0.738	0.738	0.901	-0.58	-0.39
P19	20.061	19	0.391	1.003	1.109	0.12	0.62
P20	32.929	19	0.024	1.646	1.537	1.60	2.31
P21	30.888	19	0.042	1.544	1.433	1.29	1.80
P22	21.309	19	0.320	1.065	1.007	0.30	0.13
P23	18.443	19	0.493	0.922	0.979	-0.14	-0.05
P24	14.923	19	0.727	0.746	0.885	-0.64	-0.51
P25	16.327	19	0.635	0.816	0.948	-0.48	-0.22
P26	11.394	19	0.910	0.570	0.689	-1.13	-1.50
P27	14.274	19	0.768	0.714	0.829	-0.84	-0.88
P28	18.220	19	0.508	0.911	0.861	0.03	-0.39
P29	30.731	19	0.043	1.537	1.551	1.50	2.57
P30	14.914	19	0.728	0.746	0.876	-0.66	-0.58
P31	24.267	19	0.186	1.213	0.987	0.68	0.00
P32	15.056	19	0.719	0.753	0.803	-0.70	-1.04
P33	20.937	19	0.340	1.047	1.049	0.25	0.32
P34	30.456	19	0.046	1.523	1.299	1.11	1.16
P35	14.575	19	0.749	0.729	0.837	-0.80	-0.85
P36	42.001	19	0.002	2.100	1.648	2.46	2.78
P37	22.817	19	0.246	1.141	1.154	0.52	0.84
P38	12.291	19	0.873	0.615	0.713	-1.25	-1.63
P39	17.955	19	0.525	0.898	1.021	-0.17	0.17
P40	16.536	19	0.621	0.827	0.898	-0.45	-0.50
P41	15.517	19	0.689	0.776	0.910	-0.56	-0.40
P42	26.405	19	0.119	1.320	1.241	0.99	1.25
P43	14.009	19	0.783	0.700	0.812	-0.89	-0.99
P44	19.865	19	0.403	0.993	1.113	0.15	0.50
P45	13.087	19	0.834	0.654	0.847	-0.68	-0.56
P46	18.746	19	0.473	0.937	1.046	-0.08	0.30
P47	20.727	19	0.352	1.036	0.951	0.22	-0.17
P48	34.931	19	0.014	1.747	1.105	1.24	0.44
P49	21.056	19	0.334	1.053	0.949	0.26	-0.18
P50	9.528	19	0.964	0.476	0.654	-0.97	-1.24
P51	14.220	19	0.771	0.711	0.821	-0.78	-0.88
P52	12.291	19	0.873	0.615	0.713	-1.25	-1.63
P53	13.600	19	0.806	0.680	0.791	-0.89	-1.06
P54	31.076	19	0.040	1.554	1.078	1.30	0.42
P55	32.343	19	0.029	1.617	1.171	1.23	0.71
P56	10.065	19	0.951	0.503	0.634	-1.09	-1.53
P57	16.539	19	0.621	0.827	0.953	-0.32	-0.14
P58	13.034	19	0.837	0.652	0.757	-1.08	-1.32
P59	19.288	19	0.439	0.964	1.000	0.08	0.08
P60	17.330	19	0.567	0.867	0.975	-0.31	-0.07
P61	20.120	19	0.387	1.006	0.942	0.14	-0.22
P62	16.492	19	0.624	0.825	0.846	-0.39	-0.72
P63	34.165	19	0.018	1.708	1.223	1.88	1.16

Peserta	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P64	16.539	19	0.621	0.827	0.953	-0.32	-0.14
P65	14.026	19	0.782	0.701	0.812	-0.69	-0.84
P66	12.039	19	0.884	0.602	0.701	-1.28	-1.68
P67	22.412	19	0.264	1.121	1.135	0.44	0.69
P68	15.276	19	0.705	0.764	0.881	-0.60	-0.55
P69	23.764	19	0.205	1.188	1.261	0.66	1.36
P70	12.925	19	0.842	0.646	0.748	-0.86	-1.18
P71	20.028	19	0.393	1.001	0.999	0.11	0.06
P72	17.817	19	0.535	0.891	1.028	-0.19	0.20
P73	43.315	19	0.001	2.166	1.158	2.31	0.75
P74	17.606	19	0.549	0.880	0.969	-0.26	-0.10
P75	21.770	19	0.296	1.088	1.140	0.37	0.76
P76	20.772	19	0.350	1.039	0.991	0.23	0.02
P77	22.514	19	0.259	1.126	1.185	0.43	0.84
P78	19.259	19	0.440	0.963	1.104	0.00	0.60
P79	20.772	19	0.350	1.039	0.991	0.23	0.02
P80	30.904	19	0.041	1.545	1.336	1.12	1.25
P81	13.185	19	0.829	0.659	0.993	-0.33	0.10
P82	21.352	19	0.318	1.068	0.952	0.30	-0.14
P83	16.804	19	0.603	0.840	0.963	-0.34	-0.11
P84	28.059	19	0.082	1.403	1.190	1.00	0.86
P85	19.778	19	0.408	0.989	1.058	0.11	0.33
P86	16.832	19	0.601	0.842	1.013	-0.20	0.13
P87	15.053	19	0.719	0.753	0.966	-0.29	-0.02
P88	13.238	19	0.826	0.662	0.761	-1.05	-1.32
P89	14.558	19	0.750	0.728	0.862	-0.61	-0.58
P90	13.040	19	0.836	0.652	0.754	-1.09	-1.37
P91	16.090	19	0.651	0.805	0.853	-0.54	-0.77
P92	13.179	19	0.829	0.659	0.770	-0.82	-1.05
P93	13.238	19	0.826	0.662	0.761	-1.05	-1.32
P94	17.810	19	0.535	0.891	0.921	-0.19	-0.32
P95	17.441	19	0.560	0.872	1.011	-0.25	0.12
P96	20.444	19	0.368	1.022	1.069	0.19	0.37
P97	21.290	19	0.321	1.064	1.139	0.29	0.61
P98	11.265	19	0.915	0.563	0.647	-1.47	-2.09
P99	14.558	19	0.750	0.728	0.862	-0.61	-0.58
P100	18.175	19	0.511	0.909	0.971	-0.17	-0.09
P101	11.126	19	0.920	0.556	0.699	-0.97	-1.25
P102	17.145	19	0.580	0.857	0.943	-0.34	-0.24
P103	20.324	19	0.375	1.016	1.069	0.19	0.35
P104	26.689	19	0.112	1.334	1.289	0.88	1.27
P105	14.439	19	0.758	0.722	0.845	-0.81	-0.79
P106	15.871	19	0.666	0.794	0.895	0.04	-0.10
P107	24.019	19	0.195	1.201	1.132	0.68	0.73
P108	20.777	19	0.349	1.039	1.125	0.23	0.69
P109	25.349	19	0.149	1.267	0.974	0.74	-0.04
P110	30.680	19	0.044	1.534	1.385	1.38	1.74
P111	15.360	19	0.699	0.768	0.950	-0.26	-0.07
P112	14.098	19	0.778	0.705	0.849	-0.54	-0.55
P113	15.814	19	0.670	0.791	0.951	-0.32	-0.11
P114	17.214	19	0.575	0.861	0.946	-0.29	-0.21

Peserta	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P115	16.733	19	0.608	0.837	0.992	-0.36	0.03
P116	16.733	19	0.608	0.837	0.992	-0.36	0.03
P117	21.490	19	0.310	1.075	1.109	0.33	0.63
P118	15.107	19	0.716	0.755	0.875	-0.70	-0.63
P119	17.087	19	0.584	0.854	0.983	-0.08	0.04
P120	15.107	19	0.716	0.755	0.875	-0.70	-0.63
P121	18.085	19	0.517	0.904	1.081	0.01	0.37
P122	19.379	19	0.433	0.969	1.074	0.04	0.42
P123	15.335	19	0.701	0.767	0.876	-0.66	-0.63
P124	18.751	19	0.473	0.938	1.039	-0.08	0.27
P125	12.181	19	0.878	0.609	0.761	-0.62	-0.78
P126	17.452	19	0.559	0.873	0.962	-0.12	-0.07
P127	16.803	19	0.603	0.840	0.966	-0.20	-0.05
P128	19.692	19	0.413	0.985	0.986	0.06	-0.01
P129	16.720	19	0.609	0.836	0.892	-0.43	-0.54
P130	13.197	19	0.828	0.660	0.782	-0.94	-1.07
P131	20.679	19	0.355	1.034	1.124	0.21	0.69
P132	13.742	19	0.799	0.687	0.773	-0.98	-1.26
P133	21.016	19	0.336	1.051	1.014	0.26	0.13
P134	16.121	19	0.649	0.806	0.922	-0.46	-0.33
P135	21.885	19	0.290	1.094	1.195	0.36	0.90
P136	16.973	19	0.592	0.849	0.958	-0.37	-0.16
P137	20.679	19	0.355	1.034	1.124	0.21	0.69
P138	20.224	19	0.381	1.011	1.096	0.15	0.53
P139	19.882	19	0.402	0.994	1.094	0.10	0.52
P140	17.402	19	0.563	0.870	0.982	-0.26	-0.02
P141	8.050	19	0.986	0.403	0.621	-0.90	-1.12
P142	11.619	19	0.901	0.581	0.751	-0.49	-0.65
P143	12.554	19	0.861	0.628	0.748	-1.05	-1.27
P144	23.471	19	0.217	1.174	1.241	0.62	1.27
P145	20.725	19	0.352	1.036	1.106	0.23	0.49
P146	23.410	19	0.220	1.171	1.125	0.57	0.67
P147	19.767	19	0.409	0.988	1.100	0.11	0.51
P148	20.224	19	0.381	1.011	1.096	0.15	0.53
P149	16.721	19	0.609	0.836	0.946	-0.43	-0.24
P150	16.121	19	0.649	0.806	0.922	-0.46	-0.33
P151	12.451	19	0.865	0.623	0.715	-1.21	-1.62
P152	17.021	19	0.588	0.851	0.855	-0.17	-0.52
P153	25.257	19	0.152	1.263	1.205	0.73	0.94

Dari 153 peserta tes ada 10 peserta tes yang mengalami respons yang menyimpang dari model. Hal ini terlihat kesepuluh peserta tes tersebut tidak memenuhi sebanyak dua (*p value* dan *outfit MSQ*) dari tiga kriteria person fit. Bahkan tiga peserta (P10, P36, dan P73) tidak memenuhi seluruh kriteria *person fit*. Sedang tiga peserta (P63, P80 dan P110) mempunyai peluang penerimaan *H₀* (kecocokan model) lebih kecil dari 0.05 ($p < 0.05$). Daftar peserta tes tersebut dapat disajikan dalam Tabel 4.8

Tabel 4.8 Peserta tes yang memiliki respons menyimpang (aberrant response)

Peserta	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P1	30.341	19	0.048	1.517	1.356	1.45	1.73
P10	43.579	19	0.001	2.179	1.640	2.32	2.50

Peserta	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P20	32.929	19	0.024	1.646	1.537	1.60	2.31
P21	30.888	19	0.042	1.544	1.433	1.29	1.80
P29	30.731	19	0.043	1.537	1.551	1.50	2.57
P36	42.001	19	0.002	2.100	1.648	2.46	2.78
P48	34.931	19	0.014	1.747	1.105	1.24	0.44
P54	31.076	19	0.040	1.554	1.078	1.30	0.42
P55	32.343	19	0.029	1.617	1.171	1.23	0.71
P73	43.315	19	0.001	2.166	1.158	2.31	0.75

Mendasari data di atas, dapat disimpulkan bahwa ada 3 siswa yang memenuhi 3 kriteria *aberrant response* yaitu siswa nomor P10, P36 dan P73 sehingga 98% respons peserta tes dinyatakan wajar (tidak mengalami penyimpangan) dalam mengerjakan instrumen asesmen HOTS mata pelajaran IPAS sedangkan ada 2% respons mengalami penyimpangan. Besarnya prosentase peserta tes yang memiliki respons yang wajar sesuai model ini bisa menjadi landasan bahwa tes cukup memenuhi validitas substantif.

Tes yang dibangun dalam paradigma satu dimensi haruslah benar-benar memiliki satu dimensi sehingga hasil pengukuran yang diperolehnya dapat memiliki makna. Prinsip pengujian unidimensi terlebih dahulu dinyatakan dengan hipotesis nol yang menyatakan bahwa nilai *eigenvalue* yang kedua tidak lebih besar dari nilai *eigenvalue* yang pertama dengan hipotesis alternatif bahwa nilai *eigenvalue* yang kedua lebih besar dari nilai *eigenvalue* yang pertama. Hasil analisis uji unidimensi dengan program R Menggunakan paket *ltm* dapat dilihat pada Tabel 4.9 sedangkan hasil analisis kurvanya dapat dilihat pada Gambar 4.6.

Tabel 4.9 Hasil Uji Unidimensi Instrumen Butir-Butir Instrumen HOTS IPAS

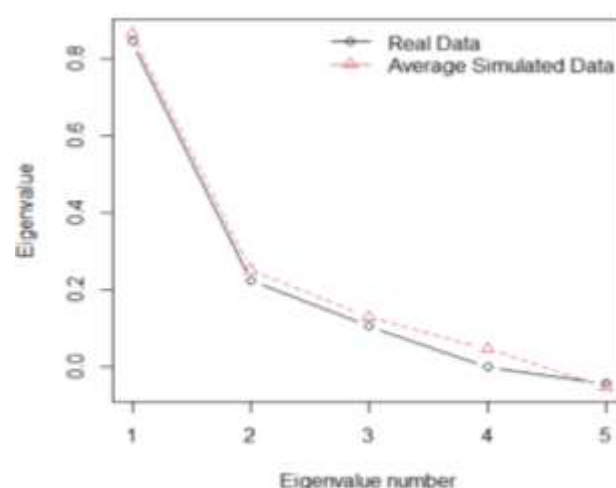
Alternative hypothesis: the second eigenvalue of the observed data is substantially larger than the second eigenvalue of data under the assumed IRT model

Second eigenvalue in the observed data: 0.2254

Average of second eigenvalues in Monte Carlo samples: 0.2528

Monte Carlo samples: 100

p-value: 0.6436



Gambar 4.6 Grafik Analisis Uji dimensionalitas Instrumen Asesmen HOTS IPAS

Terlihat bahwa peluang uji unidimensi yang dihasilkan sebesar 0.6436, suatu nilai yang lebih besar dari 0.05 sehingga dapat dinyatakan bahwa H_0 diterima. Bila H_0 diterima memiliki arti

nilai *eigenvalue* yang kedua dan seterusnya lebih kecil dari nilai *eigenvalue* yang pertama. Kondisi demikian dapat dinyatakan bahwa tes hanya mengandung satu dimensi.

Selanjutnya untuk melakukan uji invariansi pengukuran menggunakan uji *Anderson LR test*. Uji ini digunakan untuk mengetahui konsistensi estimasi parameter pemodelan Rasch. Kondisi yang ideal pada pemodelan Rasch, terjadi bila estimasi parameter tingkat kesukaran item konsisten (*invariant*) walaupun didapat dari sampel yang terdiri dari subkelompok populasi mana saja selama menerapkan pemodelan Rasch. Hasil analisis *Anderson LR test* dapat dilihat pada Tabel 4.10 Dari hasil analisis tersebut didapat nilai *p value* sebesar 0.404 artinya menerima *Ho* sehingga dapat disimpulkan bahwa estimasi parameter bersifat invarian.

Tabel 4.10 Uji Invariansi Pengukuran Menggunakan Uji Anderson LR test

Andersen LR-test: LR-value: 18.808 Chi-square df: 18 p-value: 0.404

Validitas konstruk aspek eksternal digunakan untuk mengetahui sejauh mana hasil tes didukung oleh pengukuran lain (yang mengukur domain sama atau sejenis) sehingga bisa dilihat apakah mempunyai hubungan yang kuat atau tidak. Idealnya peneliti memiliki data tes lain yang lebih akurat seperti halnya tes prestasi belajar matematika yang terstandarisasi seperti soal Asesmen Kompetensi Minimum. Dapat dimaknai bahwa uji validitas konstruk eksternal pada dasarnya adalah evaluasi terhadap suatu instrumen yang telah dikembangkan.

Salah satu pendekatan untuk mengetahui validitas konstruk aspek eksternal dalam penelitian ini adalah menggunakan informasi *Person Separation reliability* atau Separasi Person. Separasi Person digunakan untuk mengklasifikasikan orang berdasarkan informasi yang didapat dari tes. Separasi orang yang rendah (kurang dari 2) dengan sampel orang yang relevan menyiratkan bahwa instrumen mungkin tidak cukup sensitif untuk membedakan antara berkinerja tinggi dan rendah. Artinya masih dibutuhkan lebih banyak item untuk mengukurnya. Hasil analisis separasi Person menggunakan paket *eRm* dapat dilihat pada Tabel 4.11

Tabel 4.11 Uji Person Separation reliability Pada Instrumen Butir-Butir Instrumen Asesmen HOTS IPAS

Separation Reliability: 0.4427 Observed Variance: 0.4947 (Squared Standard Deviation) Mean Square Measurement Error: 0.2757 (Model Error Variance)

Berdasarkan tabel di atas nilai *Person Separation reliability* sebesar 0.4427. Menurut Sumintono & Widhiarso (2015) nilai separasi Person dapat ditentukan sebagai berikut:

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

Dengan demikian nilai separasi Person untuk tes tersebut adalah 0.9236. Dari nilai separasi Person tersebut dapat diketahui bahwa klasifikasi peserta tes yang didapat hanya 1 (pembulatan dari 0.9236). Artinya instrumen yang telah dibuat hanya membedakan peserta tes dalam dua kategori saja yaitu tuntas dan tidak tuntas. Konsekuensinya bahwa hasil tes ini hanya membedakan peserta tes menjadi dua kelompok yaitu peserta tes yang telah memiliki kecukupan minimal dan belum memiliki kecukupan minimal *High Order Thinking Skills*. Informasi ini dapat ditindaklanjuti dalam penentuan batas ketuntasan tes *High Order Thinking Skills* di SD wilayah Dabin 4 Gugus Ki Hajar Dewantara.

Aspek konsekuensi dalam validitas konstruk pada implikasi nilai interpretasi skor sebagai sumber tindakan. Bukti mengenai aspek konsekuensial validitas juga membahas konsekuensi aktual dan potensial dari pengujian dan penggunaan skor, terutama dalam hal sumber-sumber ketidakabsahan seperti bias, keadilan, dan keadilan distributif. Berkaitan dengan hal tersebut

pengukuran *High Order Thinking Skills* di SD Dabin 4 Gugus Ki Hajar Dewantara perlu dideteksi adanya bias tes.

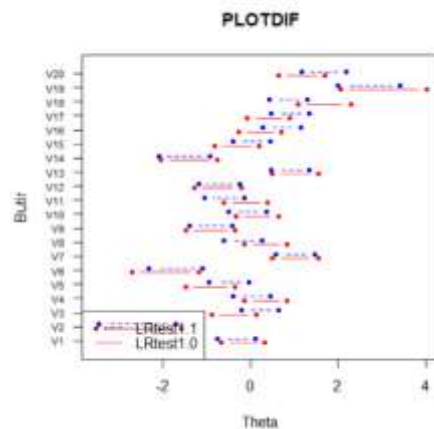
Dalam pemodelan Rasch dengan paket *eRm*, pendeteksian bias butir dapat didekati dengan menentukan butir-butir yang mengalami *differential item functioning* (DIF) Menggunakan *Wald Test*. DIF berkaitan dengan estimasi parameter butir yang berbeda pada subpopulasi yang berbeda, dalam hal ini peserta tes dibedakan berdasarkan jenis kelaminnya. Jika suatu butir dianggap lebih sukar atau lebih mudah oleh peserta tes laki-laki dibanding peserta tes perempuan atau sebaliknya, maka butir tersebut mengandung DIF. DIF atau disebut juga bias eksternal butir bukanlah justifikasi terjadinya bias butir karena untuk mengetahui ada tidaknya bias harus dilakukan kajian kualitatif mendalam lagi berkaitan penyebab munculnya DIF. Namun demikian munculnya DIF dapat menjadi petunjuk kemungkinan adanya bias.

Tabel 4.12 Hasil Analisis Dengan Wald Test Pada Tes High Order Thinking Skills

Butir	<i>z-statistic</i>	<i>p-value</i>
1	-0.451	0.652
2	-0.042	0.967
3	1.762	0.078
4	-0.962	0.336
5	1.165	0.244
6	0.456	0.648
7	-0.016	0.987
8	-1.549	0.121
9	0.037	0.970
10	-0.678	0.498
11	-1.424	0.154
12	0.096	0.924
13	-0.317	0.751
14	-0.248	0.804
15	0.988	0.323
16	1.461	0.144
17	1.475	0.140
18	-2.178	0.029
19	-0.547	0.584
20	1.337	0.181

Tabel 4.12 Daftar Butir Tes Terindikasi DIF Berdasarkan Jenis Kelamin Pada Taraf Signifikansi 0.01

Butir	<i>z-statistic</i>	<i>p-value</i>
18	-2.178	0.029



Gambar 4.7 Deskripsi DIF Pada Butir Tes Kemampuan HOTS IPAS

Kriteria secara statistik dengan *Wald test*, butir yang mengalami DIF adalah yang memiliki *p-value* kurang dari 0.01 (bila menggunakan taraf signifikansi 0.01).

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan pada bab sebelumnya, peneliti dapat menyimpulkan sebagai berikut :

1. Guru dan siswa tingkat Sekolah Dasar sangat membutuhkan pengembangan asesmen HOTS mata pelajaran IPAS.
2. Desain asesmen HOTS mata pelajaran IPAS menggunakan desain penelitian dan pengembangan (*Riset and development*) model ADDIE pada tahap analisis, desain dan pengembangan.
3. Hasil uji validasi isi dan validasi psikometri menunjukkan hasil yang valid sehingga instrumen asesmen HOTS IPAS dapat diterapkan pada tingkat Sekolah Dasar.
4. Hasil uji validasi konstruk (aspek isi, substantif, struktural, eksternal, konsekuensial) dengan pemodelan *Rasch* menunjukkan hasil yang valid sehingga instrumen asesmen HOTS IPAS dapat diterapkan pada tingkat Sekolah Dasar.

UCAPAN TERIMA KASIH

Penulis berterimakasih kepada semua pihak yang secara langsung dan tidak langsung memberikan kontribusi dalam penyelesaian tesis ini. Secara khusus penulis menyampaikan terimakasih kepada :

1. Dr. Taufiqulloh, M.Hum. Selaku Rektor Universitas Pancasakti Tegal atas motivasi, arahan, dan dukungan moralnya.
2. Prof. Dr. Sitti Hartinah, DS, M.M. Selaku Direktur Program Pascasarjana Universitas Pancasakti Tegal dan Pembimbing I, yang telah mensupport dan memberikan motivasi dalam penyelesaian tesis ini.
3. Dr. Suriswo, M.Pd. Selaku Ketua Program Magister Pedagogi Pascasarjana Universitas Pancasakti Tegal atas bimbingan, arahan, dan waktu yang telah diluangkan kepada penulis untuk berdiskusi selama penyusunan tesis.
4. Prof. Dr. Purwo Susongko, M.Pd Selaku Pembimbing II atas bimbingan, arahan, dan waktu yang telah diluangkan kepada penulis untuk berdiskusi selama penyusunan tesis.
5. Seluruh Dosen Program Studi Magister Pedagogi Pascasarjana Universitas Pancasakti Tegal yang telah memberikan arahan dan bimbingan untuk mendalami Ilmu Pendidikan beserta staffnya.
6. Rekan-rekan guru di Dabin 4 Gugus Ki Hajar Dewantara KWK Dikbud Kecamatan Talang yang telah bersedia meluangkan waktunya untuk berpartisipasi sebagai informan dalam penelitian ini yang tidak dapat saya sebutkan satu persatu.

7. Semua pihak yang tidak bisa saya sebutkan satu persatu, yang telah membantu dan mendoakan.

DAFTAR PUSTAKA

- Aldila, F. T., Darmaji, D., & Kurniawan, D. A. (2022). Analisis Respon Pengguna terhadap Penerapan Web-based Assessment pada Penilaian Sikap Siswa terhadap Mata Pelajaran IPA dan Nilai-nilai Pendidikan Karakter. *Edukatif: Jurnal Ilmu Pendidikan*, 4(1), 1253-1262.
- Anggito, A., & Setiawan, J. (2018). *Metodologi penelitian kualitatif*. CV Jejak (Jejak Publisher).
- Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan edisi 3*. Bumi aksara.
- Asrial, A., Syahrial, S., Maison, M., Kurniawan, D. A., & Piyana, S. O. (2020). Ethnoconstructivism e-module to improve perception, interest, and motivation of students in Class V Elementary School. *JPI (Jurnal Pendidikan Indonesia)*, 9(1), 30-41.
- Asrijanty, A. (2020). Asesmen Kompetensi Minimum (AKM) dan implikasinya pada pembelajaran.
- Astawayasa, K. G., Widana, I. W., & Adi, I. N. R. (2022). Pengembangan asesment HOTS mata pelajaran matematika sekolah dasar. *Jurnal Ilmiah Pendidikan Citra Bakti*, 9(1), 129-140.
- Azizah, A., & Wahyuningsih, S. (2020). Penggunaan model RASCH untuk analisis instrumen tes pada mata kuliah matematika aktuaria. *Jurnal Pendidikan Matematika (JUPITEK)*, 3(1), 45-50.
- Azzahri, C. K., Widjanarko, D., & Sudana, I. M. (2017). Pengembangan Instrumen Penilaian Praktik Rias Pengantin Jogja Paes Ageng pada Mata Kuliah Rias Pengantin Jawa. *Journal of Vocational and Career Education*, 2(1).
- Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal penelitian dan evaluasi pendidikan*, 22(2), 219-230.
- Desiriah, E., & Setyarsih, W. (2021). Tinjauan literatur pengembangan instrumen penilaian kemampuan berpikir tingkat tinggi (HOTS) fisika di sma. *ORBITA: Jurnal Pendidikan dan Ilmu Fisika*, 7(1), 79-89.
- Ennis, R. H. (2015). Critical thinking: A streamlined conception. In *The Palgrave handbook of critical thinking in higher education* (pp. 31-47). New York: Palgrave Macmillan US.
- Fanani, M. Z. (2018). Strategi pengembangan soal hots pada kurikulum 2013. *Edudeena: Journal of Islamic Religious Education*, 2(1), 57-76.
- Freeman, E. W., Sammel, M. D., Boorman, D. W., & Zhang, R. (2014). Longitudinal pattern of depressive symptoms around natural menopause. *JAMA psychiatry*, 71(1), 36-43.
- Hamidah, N., Haryani, S., & Wardani, S. (2018). Efektivitas lembar kerja peserta didik berbasis inkuiri terbimbing untuk meningkatkan hasil belajar siswa. *Jurnal Inovasi Pendidikan Kimia*, 12(2).
- Inasari, L., Lidinillah, D. A. M., & Prehanto, A. (2023). Pengembangan instrumen tes computational thinking Siswa Sekolah Dasar melalui analisis RASCH model. *COLLASE (Creative of Learning Students Elementary Education)*, 6(1), 102-110.
- Iskandar, S. M. (2016). Pendekatan keterampilan metakognitif dalam pembelajaran sains di kelas. *Erudio Journal of Educational Innovation*, 2(2), 13-20.
- Juhaeni, J., Wiji, S., Wadud, A. J., Saputra, H., Azizah, I. N., & Safaruddin, S. (2022). Pengaruh Media Pembelajaran Teka Teki Silang Terhadap Hasil Belajar IPA Materi Perkembangbiakan Tumbuhan. *Journal of Instructional and Development Researches*, 2(6), 241-247.
- Kamus Besar Bahasa Indonesia (KBBI) online, <https://kbbi.kemdikbud.go.id/> diakses tanggal 4 Januari 2024.
- Kemendikbubristek Nomor 262/M/2022 Tentang Pedoman Penerapan Kurikulum Dalam Rangka Pemulihan Pembelajaran.
- Keputusan Kepala BSKAP Nomor 034/H/KR/2022 tentang Satuan Pendidikan Pelaksana Implementasi Kurikulum Merdeka.
- Khadijah, K., & Amelia, N. (2020). Asesmen perkembangan kognitif anak usia 5-6 tahun. *Al-athfaal: jurnal ilmiah pendidikan anak usia dini*, 3(1), 69-82.

- Khayati, D. N., & Raharjo, R. (2020). Pengembangan Instrumen Tes Berbasis Literasi Sains untuk Memetakan Critical Thinking dan Practical Skills Siswa pada Materi Sistem Peredaran Darah Kelas XI SMA. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, 9(3), 433-442.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218.
- Kunandar, A. (2014). Model literasi media pada anak dalam mencegah konflik sosial. *Profetik: Jurnal Komunikasi*, 7(1).
- Kusuma, A. S., & Nurmawanti, I. (2023). Pengembangan Soal-Soal Literasi dan Numerasi Berbasis High Order Thinking Skills (HOTS) untuk Siswa Sekolah Dasar (SD). *Jurnal Ilmiah Profesi Pendidikan*, 8(1), 516-523.
- Liana, N., Suana, W., Sesunan, F., & Abdurrahman, A. (2018). Pengembangan Soal Tes Berpikir Tingkat Tinggi Materi Fluida Untuk SMA. *Journal of Komodo Science Education*, 1(1), 66-78.
- Lie, A., Tamah, S. M., Gozali, I., & Triwidayati, K. R. (2020). *Mengembangkan keterampilan berpikir tingkat tinggi*. PT Kanisius.
- Marudut, M. R. H., Bachtar, I. G., Kadir, K., & Isha, V. (2020). Peningkatan Kemampuan Berpikir Kritis dalam Pembelajaran IPA melalui Pendekatan Keterampilan Proses. *Jurnal Basicedu*, 4(3), 577-585.
- Nisrokha, N. (2020). Test Terstandarisasi. *Madaniyah*, 10(1), 15-36.
- Pakaya, W. C., Sutadji, E., Dina, L. N. A. B., Rahma, F. I., Mashfufah, A., & Ayu, I. R. (2023). *Metode Penelitian Pendidikan*. Nawa Litera Publishing.
- Permendikbud nomor 22 Tahun 2020 Tentang Kurikulum Masa Peralihan (Kurikulum Merdeka).
- PP No.19 Tahun 2005 tentang Standar Nasional Pendidikan
- Rahayu, L. P., Nugroho, A. S., Santoso, M., & Widodo, S. (2018). Pengembangan Soal Matematika HOTS (Higher Order Thinking Skills) Kelas X Berdasarkan Triple Theory. *Repository Publikasi Ilmiah*, 117-125.
- Retnawati, H. (2017, July). Membuktikan validitas instrumen penelitian. In *Makalah Disajikan Pada Kegiatan Workshop Teknik Analisis Data Fakultas Ekonomi Dan Bisnis IAIN Batusangkar Di Rocky Hotel Bukittinggi, 25 Juli 2017* (pp. 1-16).
- Retnowati, T. H., Mardapi, D., Kartowagiran, B., & Suranto, S. (2017). Model evaluasi kinerja dosen: pengembangan instrumen untuk mengevaluasi kinerja dosen. *Jurnal penelitian dan evaluasi Pendidikan*, 21(2), 206-214.
- Safaruddin, S., Nurhastuti, N., Fatmawati, F., & Silitonga, E. C. (2018). Efektivitas Metode Survey, Question, Read, Recite, Reflect, Review dalam Meningkatkan Keterampilan Membaca Pemahaman untuk Anak Berkesulitan Belajar. *Jurnal Pendidikan Kebutuhan Khusus*, 2(2), 1-5.
- Sahlani, L., & Agung, B. (2020). Asesmen pembelajaran berbasis google form pada mata pelajaran sejarah kebudayaan islam di MAN 2 Bandung. *AL-IBANAH*, 5(1), 1-27.
- Sanjaya, H. W. (2015). *Penelitian Pendidikan: Metode, Pendekatan, dan Jenis*. Kencana.
- Saputra, H. (2016). *Pengembangan mutu pendidikan menuju era global: Penguatan mutu pembelajaran dengan penerapan hots (high order thinking skills)*. Smile's.
- Saputro, B. (2017). *Manajemen penelitian pengembangan (research & development) bagi penyusunan tesis dan disertasi*. Aswaja Presindo.
- Sari, A., & Yuniati, S. (2018). Penerapan pendekatan realistic mathematics education (RME) terhadap kemampuan pemahaman konsep matematis. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 2(2), 71-80.
- Septiana, A. N., & Winangun, I. M. A. (2023). Analisis Kritis Materi IPS dalam Pembelajaran IPAS Kurikulum Merdeka di Sekolah Dasar. *Widyaguna: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 1(1), 43-54.
- Subay, R. (2020). "Pengembangan Asesmen High Order Thinking Skills (HOTS) siswa pada pembelajaran matematika kelas VII berbasis model Rasch".
- Sudijono, A. (2016). *Pengantar Evaluasi Pendidikan*. Edisi Pertama
- Sugimin (2022). "Model Asesmen HOTS mata pelajaran IPAS pada siswa SMK pusat keunggulan di SMK Negeri 1 Adiwerna".

- Sugiyono, A. B., & WIJAYANTI, P. (2018). Representasi Matematis Siswa Dalam Menyelesaikan Soal Model PISA Matematika Ditinjau dari Perbedaan Jenis Kelamin. *MATHEdunesa*, 7(3), 619-623.
- Sugiyono, S., Sutarman, S., & Rochmadi, T. (2019). Pengembangan sistem computer based test (CBT) tingkat sekolah. *Indonesian Journal of Business Intelligence (IJUBI)*, 2(1), 1-8.
- Suprijanto, E., & Arikunto, S. (2016). Efektivitas pengelolaan kegiatan kelompok kerja guru (KKG) di Kecamatan Rembang, Purbalingga, Jawa Tengah. *Jurnal Penelitian Ilmu Pendidikan*, 9(2), 141-151.
- Susongko, P. (2019). Aplikasi Model Rasch Dalam Pengukuran Pendidikan Berbasis Program R.
- Susongko, P., Kusuma, M., Arfiani, Y (2019). "Model Asesmen Literasi Sains Siswa Berbasis IPA Terpadu Dengan Pemodelan Rasch Untuk Peningkatan Kompetensi Lulusan SMA Program Matematika Dan Ilmu Alam (MIPA)".
- Susongko, P., Widiatmo, H., Kusuma, M., & Afiani, Y. (2019). Development of integrated science-based science literacy skills instruments using the Rasch model. *Unnes Science Education Journal*, 8(3).
- Susongko, P., Yuenyong, C., & Zainudin, A. (2022). Buddhist critical thinking assessment using Rasch model. *Kasetsart Journal of Social Sciences*, 43(2), 285-292.
- Undang-Undang Nomor 20 Tahun 2003 Tentang Sistem Pendidikan Nasional.
- Widiana, I. W. (2016). Pengembangan asesmen proyek dalam pembelajaran ipa di sekolah dasar. *JPI (Jurnal Pendidikan Indonesia)*, 5(2), 147-157.
- Wijayanti, I., & Ekantini, A. (2023). Implementasi Kurikulum Merdeka pada Pembelajaran IPAS MI/SD. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, 8(2), 2100-2112.
- Wilson, L. O. (2016). Anderson and Krathwohl–Bloom’s taxonomy revised. *Understanding the new version of Bloom's taxonomy*.
- Wilson, L. O. (2016). Anderson and Krathwohl–Bloom’s taxonomy revised. *Understanding the new version of Bloom's taxonomy*.
- Wirda, M. A., Berutu, N., & Rahmad, R. (2017). Pengembangan Tes Standar Berasis Teknologi Informasi Di Jurusan Pendidikan Geografi. *Tunas Geografi*, 6(2), 101-115.
- Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel Rasch modeling of two-tier multiple choice test: A case study using Lawson’s classroom test of scientific reasoning. *Physical Review Physics Education Research*, 14(2), 020104.
- Yani, J. A. Sugiyono. 2017. Metode Penelitian Kuantitatif, Kualitatif, Dan R&D. Bandung: Alfabeta.
- Ferrari, JR, Jhonson, JL, & McCown, WG (1995). Procrastination And Task Avoidance: Theory, Research & Treatment. New York: Plenum Press.
- Yudistira P, Chandra. Diktat Kuliah Psikometri. Fakultas Psikologi Universitas.
- Yanitsky, O. N. (2017). Metabolism as a master frame for globalization analysis.