Pengembangan Penilaian untuk Mengukur Kemampuan Literasi Membaca Peserta Didik SMP dengan Standar PISA 2018

Nunung Chasanah^{1✉}, Purwo Susongko², Suriswo³
(1,2,3) Pedagogi, Universitas Pancasakti Tegal

✉ Corresponding author

[\[nunungchasanah12@guru.smp.belajar.id\]](mailto:nunungchasanah12@guru.smp.belajar.id)

Abstrak

Kemampuan Literasi membaca peserta didik di UPTD SPF SMP N 10 Tegal masih rendah. Penelitian ini bertujuan untuk mengembangkan instrumen tes untuk meningkatkan kemampuan literasi membaca peserta didik SMP berstandar PISA. Penelitian ini merupakan Penelitian Pengembangan atau *Research and Development* (R & D) dan dilaksanakan di UPTD SPF SMP N 10 Tegal. Peneliti menganalisis kebutuhan, konstruksi tes, validitas aspek isi, aspek psikometri butir-butir, dan validitas kontrak tes literasi membaca untuk peserta didik UPTD SPF SMP N 10 Tegal. Kesimpulan penelitian menunjukkan (1) adanya kebutuhan pengembangan tes literasi membaca bagi Guru dan Peserta Didik di UPTD SPF SMP N 10 Tegal; (2) Konstruksi tes telah mengacu pada indikator standar PISA; (3) Instrumen tes telah memenuhi validitas dari aspek isi; (4) instrumen tes telah memenuhi validitas aspek psikometrik; dan (5) terdapat 1 butir soal tidak terpakai dan 29 butir soal layak sehingga dapat dimanfaatkan untuk mengukur kemampuan literasi membaca.

Kata Kunci: Pengembangan, Literasi Membaca, SMP, PISA

Abstract

The reading literacy ability of students at UPTD SPF SMP N 10 Tegal is still low. This research aims to develop a test instrument to improve the reading literacy skills of PISA standard junior high school students. The researcher analyzed the need, test construction, validity of content aspects, psychometric aspects of items, and validity of the reading literacy test contract for UPTD SPF SMP N 10 Tegal students. The research conclusions show (1) there is a need to develop a reading literacy test for Teachers and Students at UPTD SPF SMP N 10 Tegal; (2) The test construction refers to the PISA standard 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) there is 1 unused question item and 29 suitable questions so that it can be used to measure reading literacy skills.

Keyword: Development, Reading Literacy, Middle School, PISA

PENDAHULUAN

Pendidikan abad ke-21 tergolong kompetitif dan bergerak dengan dinamis. Untuk itu, pendidikan harus menyiapkan peserta didik agar mampu bertahan hidup dan berkontribusi aktif dengan berbagai hal. PISA (the programme for international student assessment) merupakan program untuk mengukur prestasi bagi anak usia 15 tahun pada bidang kemampuan matematika, sains dan literasi membaca. Penilaian yang dilakukan oleh PISA dilakukan tiap 3 (tiga) tahun sekali dengan fokus pada pendidikan suatu negara. Adapun negara-negara yang berpartisipasi pada penilaian PISA semenjak pertama kali dilakukan yaitu sejak tahun 2000 terus bertambah, tercatat hingga 2018 dari 41 menjadi 79 negara sebagai partisipan dalam penilaian PISA di bawah Organization for Economic Co-operation and Development (OECD, 2019b).

Literasi membaca menjadi kemampuan awal yang penting. literasi membaca diartikan sebagai wawasan, daya mampu membaca dan menulis, mengorganisasi, dan memahami segala info yang diperoleh ketika dilakukannya kegiatan membaca dan menulis, serta kemampuan menganalisis, berpendapat, dan penggunaan bahasa (Kemendikbud, 2019). Kemampuan literasi membaca telah dijadikan sebagai syarat awal saat keikutsertaan kegiatan yang berbau sosial, kultural, politis, dan ekonomis pada zaman modern. Dengan demikian, kemampuan literasi membaca akan memudahkan seseorang beradaptasi di masyarakat sosial. Pembelajaran literasi membaca di sekolah dimuat pada kurikulum di bagian Standar Isi (Permendikbud, 2018). Penilaian perlu dilakukan untuk mengetahui bagaimana penguasaan peserta didik SMP pada mapel bahasa Inggris terhadap capaian literasi membaca yang telah disesuaikan dengan standar PISA 2018 melalui evaluasi atau tes. Secara garis besar, alat evaluasi yang digunakan dapat digolongkan menjadi 2 (dua) macam, yaitu tes dan bukan tes/ non tes (Arikunto, S, 201). Hasil yang dimaksud dalam tes ini adalah sejauh mana peserta didik dapat menunjukkan deskripsi kemampuan literasi membaca dengan menjawab benar dari bacaan-bacaan yang diujikan lewat butir soal tes.

Hasil penelitian mengungkapkan bahwa instrument soal tes literasi juga layak digunakan untuk meningkatkan kemampuan literasi siswa, karena memperoleh penilaian yang valid dan reliabel (Putri, 2020). Hasil PISA mengindikasikan bahwa skor literasi membaca peserta didik Indonesia masih tergolong rendah dibandingkan dengan negara-negara lain di dunia. Pada tahun 2018, skor literasi membaca Indonesia berada di peringkat 57 dari 79 negara peserta. Hal ini menunjukkan bahwa kemampuan literasi membaca peserta didik Indonesia masih perlu ditingkatkan Rendahnya skor literasi membaca peserta didik Indonesia dapat disebabkan oleh berbagai faktor, antara lain: 1) Kurangnya minat baca pada peserta didik yaitu karena kurangnya budaya literasi di sekolah dan masyarakat, kualitas pembelajaran literasi di sekolah yang masih rendah, serta instrumen penilaian literasi membaca yang belum valid dan reliabel; 2) Kebutuhan akan instrumen penilaian literasi membaca yang valid dan reliabel dapat memberikan informasi yang akurat tentang kemampuan literasi membaca peserta didik secara keseluruhan khususnya dalam aspek kognitif.

Oleh karena itu, diperlukan pengembangan instrumen penilaian yang lebih valid dan reliabel untuk mengukur kemampuan literasi membaca pada mapel bahasa Inggris peserta didik SMP kelas VIII berstandar PISA dengan pendekatan Rasch Model.

METODE PENELITIAN

Penelitian ini merupakan Penelitian Pengembangan atau *Research and Development (R & D)* dan dilaksanakan di UPTD SPF SMP N 10 Tegal. Objek penelitian ini adalah instrument penilaian literasi membaca berstandar PISA dengan model Rasch pada Mata Pelajaran Bahasa Inggris. Ada 3 tahap penelitian yang menggunakan metode penelitian dan pengembangan. Winaryati, dkk (2021:8) merinci 3 tahap tersebut terdiri dari *Research, Development, dan Diffusion*.

Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan di UPTD SPF SMP N 10 Tegal di Jl. Kartini No. 58, Desa Mangkukusuman, Kec. Tegal Timur, Kota Tegal Propinsi Jawa Tengah. Penelitian dilaksanakan pada bulan Januari sampai Juni 2024. Tahun Pelajaran 2023/2024

Populasi, Sampel, dan Teknik Sampling

Populasi dalam penelitian ini adalah seluruh responden Kelas VIII sebanyak tujuh kelas dengan jumlah total peserta didik 224. Sampel dalam penelitian ini menggunakan teknik dengan metode non probability sampling dan teknik *Purposive Sampling*.

Teknik Pengumpulan Data

Pengumpulan Data pada penelitian ini menggunakan 2 teknik yaitu observasi, dokumentasi, validitas pakar dan perangkat tes.

Langkah-Langkah Penelitian

Pelaksanaan penelitian sesuai desain penelitian ini menggunakan pendekatan ADD (*Analysis, Design, Development*).

Analisis

Tahap Analisis merupakan kegiatan awal untuk mengetahui kebutuhan dan tujuan produk yang akan dikembangkan. Produk penelitian ini adalah instrument penilaian literasi membaca berstandar PISA pada mata pelajaran Bahasa Inggris SMP. Dalam penyusunan instrumen dilakukan

uji validitas untuk menjamin kesesuaian instrumen yang dikembangkan dengan tujuan yang ingin dicapai menggunakan indikator *PISA* berdasarkan capaian pembelajaran mata pelajaran di kurikulum merdeka.

HASIL DAN PEMBAHASAN

Analisis Kebutuhan

Model pengembangan Tes literasi membaca 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 literasi membaca pada mata pelajaran bahasa Inggris berstandar *PISA* yang terdiri dari 30 butir soal pilihan ganda.

Tahap perencanaan instrumen meliputi penentuan materi, aspek/ dimensi yang dinilai, dan indikator yang digunakan dalam pengembangan Tes Bahasa Inggris bestandar *PISA*. Materi yang menjadi bahan penyusunan instrumen pada penelitian ini merupakan materi ajar yang telah diberikan kepada Peserta Didik Kelas VIII di UPTD SPF SMP N 10 Tegal. Sedangkan dimensi atau aspek yang dinilai melalui instrumen penelitian meliputi tiga aspek dalam *PISA* yaitu menemukan informasi (*retrieve and access*), menginterpretasi / mengintegrasikan (*interpret and integrate*), evaluasi dan refleksi (*evaluate and reflect*). Indikator yang digunakan dalam penyusunan instrumen soal merupakan penyesuaian antara materi dan aspek yang akan diukur dalam penelitian ini.

Instrumen Tes diujicobakan kepada peserta didik Kelas VIII 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 literasi membaca berstandar *PISA* dan divalidasi oleh para ahli. Hasil Validasi Isi Angket Kebutuhan Instrument literasi membaca tersebut adalah sebagai berikut:

Tabel 1 Rubrik Validasi Isi Angket Kebutuhan Instrument literasi membaca

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	√		√	

Sumber : Data Primer

Hasil Analisis kebutuhan per Indikator berdasarkan Tes ujicoba yang dilakukan pada 40 guru

Tabel 2 Analisis kebutuhan per Indikator

No	Indikator	Prosentase
1	Retrieve and access (menemukan informasi)	47,00%
2	Interpret and integrate (menginterpretasi dan mengintegrasikan)	43,00%
3	Evaluate and reflect (mengevaluasi dan merefleksikan)	36,00%
4	<i>PISA</i>	48,00%

Sumber : Diolah dari data primer dengan R Program

Hasil analisis kebutuhan per Indikator ditemukan bahwa perolehan skor responden pada semua indikator literasi berada di bawah 50% atau rendah, sehingga dibutuhkan pengembangan instrumen Tes untuk meningkatkan literasi membaca peserta didik.

Analisis Materi

Pengukuran materi bahasa Inggris dalam pengembangan tes literasi membaca di kelas VIII disesuaikan dengan Capaian Pembelajaran antara lain pembahasan pada tema teks Recount, Narrative, Descriptive, Degree of Comparison yang terdapat dalam materi bahasa Inggris kelas VIII SMP. Adapun indikator yang dibahas untuk literasi membaca yaitu: (1) mengidentifikasi teks informasi dan teks fiksi dan non fiksi, (2) mengevaluasi proses kognitif yang terdiri dari mencari informasi, memahami teks (3) merefleksikan dan mengevaluasi teks pribadi, sosial budaya dan konteks ilmiah.

Adapun butir soal tes literasi membaca dalam bentuk pilihan ganda bertujuan untuk menguji pemahaman kognitif peserta didik mengenai bermacam teks bacaan berupa kontekstual bacaan yang disesuaikan dengan materi bahasa Inggris pada Kurikulum Merdeka saat ini. Agar tujuan pengukuran kemampuan literasi membaca tercapai, maka pilihan jawaban disusun secara acak dan tidak mengikuti pola.

Design

Desain pengembangan yang dilakukan dalam penulisan soal literasi membaca pada penelitian ini meliputi tahap-tahap sebagai berikut: (1) menentukan tujuan tes dan materi, untuk memastikan bahwa soal yang dibuat sesuai dengan materi yang dipelajari oleh peserta uji; (2) Menyusun kisi-kisi, disesuaikan dengan acuan penulisan soal dan indikator capaian pembelajaran yang telah dipelajari oleh peserta uji; (3) pemilihan Format tes, berupa pilihan ganda karena memungkinkan jumlah responden keseluruhan dalam waktu yang relatif singkat dan memudahkan dalam analisis hasil tes; (4) menentukan durasi tes, peneliti mengalokasikan waktu sekitar 60 menit hingga 90 menit didasarkan alokasi satu jam pelajaran di sekolah; (5) penyusunan soal tes, disesuaikan dengan kisi-kisi dan indikator tes agar dihasilkan instrumen yang berkualitas baik.

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 literasi membaca (PISA)

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 indikator dari bacaan (PISA)	√		√	
3	Kunci jawaban benar	√		√	

Sumber : Data Primer

Hasil Validasi Isi pada menunjukkan instrumen Tes bahasa Inggris yang dikembangkan untuk meningkatkan kemampuan literasi membaca 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 Literasi membaca pada mata pelajaran bahasa Inggris 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 20 Mei 2024 di UPTD SPF SMP N 10 Tegal. Hasil analisis dengan model Rasch berisi hasil analisis kecocokan butir terhadap model (*item fit*).

Tabel 4 Hasil Analisis Item Fit Instrumen Tes Literasi Peserta Didik Kelas VIII SMP N 10 Tegal

No Butir	Chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
V1	214.882	223	0.640	0.959	0.963	-0.493	-0.508
V2	221.398	223	0.518	0.988	0.994	-0.103	-0.045
V3	241.012	223	0.198	1.076	1.030	0.719	0.335
V4	213.811	223	0.659	0.995	0.979	-0.530	-0.258
V5	241.989	223	0.182	1.080	1.042	0.787	0.476
V6	225.244	223	0.445	1.006	1.014	0.095	0.210
V7	225.490	223	1.441	1.007	1.002	0.100	0.052
V8	209.600	223	0.731	0.936	0.996	-0.625	-0.353
V9	218.947	223	0.564	0.997	0.990	-0.241	-0.107
V10	245.675	223	0.142	1.097	1.056	1.052	0.693
V11	233.709	223	0.298	1.043	1.019	0.496	0.252
V12	213.118	223	0.671	0.951	0.979	-0.429	-0.196
V13	256.626	223	0.000	1.146	1.071	1.427	0.814
V14	200.921	223	0.853	0.897	0.939	-0.884	-0.571
V15	243.162	223	0.169	1.086	1.052	1.053	0.728
V16	221.132	223	0.523	0.987	1.001	-0.099	0.037
V17	194.935	223	0.913	0.870	0.930	-0.936	-0.537
V18	232.252	223	0.321	1.037	1.003	0.405	0.058
V19	211.343	223	0.702	0.943	0.984	-0.494	-0.138
V20	207.273	223	0.768	0.970	0.970	-0.654	-0.270
V21	234.161	223	0.291	1.000	1.000	0.400	0.032
V22	217.224	223	0.596	0.996	0.996	-0.259	-0.016
V23	205.648	223	0.792	0.958	0.958	-0.673	-0.365
V24	219.226	223	0.559	1.001	1.001	-0.142	-0.046
V25	221.646	223	0.513	0.990	0.990	-0.084	-0.096
V26	224.423	223	0.461	1.010	1.010	0.054	0.133
V27	215.002	223	0.637	0.983	0.983	-0.357	-0.155
V28	236.306	223	0.258	1.048	1.048	0.553	0.536
V29	222.215	223	0.502	0.998	0.998	-0.026	0.016
V30	230.503	128	0.351	1.029	1.015	0.350	0.216

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 0.01 ($p > 0.01$).

Seluruh butir secara umum dapat diterima sebagai butir yang baik kecuali butir nomor 13. Butir nomor 13 memiliki $p\text{-value} < 0.00$ dapat disimpulkan bahwa taraf signifikan 0.00 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 -8,612 sampai 3,172.

Tabel 5 Nilai Tingkat Kesukaran Butir-butir Instrumen HOTS

No Butir	Nilai Kesukaran	No Butir	Nilai Kesukaran
1	1.087	16	-1.179
2	-2.261	17	-2.608
3	-7.610	18	-3.804
4	-2.108	19	-0.264
5	-3.443	20	-4.422
6	-1.821	21	-1.637
7	-2.563	22	-0.352
8	-3.689	23	1.563
9	-4.786	24	1.064
10	-0.057	25	1.231
11	0.000	26	-1.266
12	-8.612	27	-2.768
13	-3.104	28	3.172
14	-2.802	29	1.511
15	-2.181	30	1.080

Sumber : Diolah dari data primer dengan R Program

Dari tabel di atas memperlihatkan tingkat kesulitan kesukaran paling rendah (-8.610) merupakan item tes nomor 12, sedangkan tingkat kesukaran paling tinggi (+3.172) item tes nomor 28. Tingkat kesukaran sebesar 3.172 merupakan perolehan skor kemampuan minimal -3.172 maka peserta tes diprediksi dapat mengerjakan butir dengan benar.

Berdasarkan data diatas terdapat 22 butir item tes termasuk kategori rendah, 1 item tes termasuk kategori sedang, dan 7 item tes yang termasuk kategori sukar pada pengembangan butir soal Bahasa Inggris yang peneliti susun dalam penelitian ini. Tingkat kesukaran butir soal merupakan parameter lokasi menunjukkan posisi kurva karakteristik butir dalam hubungannya dengan skala kemampuan. Parameter tingkat kesukaran digambarkan pada suatu titik jika menjawab benar sebesar 0,5. Semakin tinggi nilai parameter tingkat kesulitan yang maka dibutuhkan tingkat kemampuan yang tinggi dari responden untuk memiliki peluang 0,5 dalam menjawab soal dengan benar.

Tingkat kesukaran butir bergerak dari -2.922 hingga 0.833. 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 pengukuran butir soal literasi membaca 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 tingkatkesukaran butir untuk pengukuran yang bersifat kompetensi dapat dimulai dari -3.00 hingga 3.00, karena pada interval tersebut dapat mengukur sekitar 100% peserta tes. Dengan demikian dari hasil analisis semua butir-butir tes instrumen literasi membaca yang telah disusun, berada pada interval -3.00 hingga 3.00, karena pada interval tersebut dapat mengukur sekitar 95 % peserta tes sehingga efektif sebagai tes kompetensi.

Validitas Konstruk Aspek Subtantif

Secara kuantitatif uji respons terjadinya penyimpangan atau tidak adanya penyimpangan disebut person fit. 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 ($p > 0.01$). Tabel dibawah ini berisi hasil uji person fit dari 224 respons tes literasi membaca peserta didik SMP N 10 Tegal ada 11 peserta tes yang mengalami respon yang menyimpang dari model. Hal ini terlihat dari ke 11 peserta tes tersebut tidak memenuhi sebanyak dua (p value dan outfit MSQ) dari ketiga

kriteria person fit. Dari hasil uji person fit berikut hasil analisis validitas substantif Person fit statistic tercantum pada tabel berikut: respon 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 ($p > 0.01$). 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
P16	21.132	29	0.000	0.987	1.001	4.49	0.07
P26	19.398	29	0.000	1.015	0.994	-0.10	-0.45
P45	24.012	29	0.001	1.567	1.030	0.71	0.35
P53	23.811	29	0.003	0.979	0.979	2.53	-0.58
P60	30.656	29	0.002	0.415	1.042	0.88	0.76
P63	28.233	29	0.000	4.606	1.063	2.09	0.21
P72	25.490	29	0.001	1.707	1.002	0.14	0.12
P77	29.600	29	0.001	1.936	1.296	4.42	2.35
P85	29.387	29	0.003	0.319	0.990	-1.24	1.17
P89	31.675	29	0.000	0.980	0.995	1.05	0.69
P91	33.705	29	0.000	1.643	1.019	0.49	0.25

Sumber : Diolah dari data primer dengan R Program

Terdapat 95.08% respons peserta tes yang wajar dan sesuai dengan model, sedangkan respon peserta tes yang mengalami penyimpangan sebesar 4.91%. 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.		.561
Bartlett's Test of Sphericity	Approx. Chi-Square	10.492
	Df	10
	Sig.	.387

Sumber : Diolah dari data primer

Angka KMO *Measure of sampling Adequacy* (MSA) adalah 0.561. Hal ini menunjukkan kecukupan dari sampel. Angka KMO dan *Bartlett's test* (yang tampak pada nilai *chi-square*) sebesar 10.492 dengan nilai signifikansi 0.387. 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 0.431 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 konstruk 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 menunjukkan hasil nilai Cronbach's Alpha sebesar 0,674 sehingga dinyatakan memenuhi validasi konstruk aspek eksternal.

Pembahasan

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 instrumen 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.

Item fit berfungsi untuk mendeteksi adanya kesalahan yang terjadi selama fase penentuan standar pengukuran yang akurat pada pengembangan instrumen Tes. Hal tersebut teridentifikasi melalui parameter daya beda. Jika parameter daya beda menunjukkan hasil yang tidak baik, maka *item fit statistic* dapat mengidentifikasi masalah sehingga dapat dilakukan perbaikan instrumen Tes. Sedangkan *person fit*, berfungsi untuk mengidentifikasi ada atau tidaknya penyimpangan pola respon dari peserta tes. Penyimpangan tersebut dapat terjadi karena adanya perilaku curang dalam mengerjakan soal, spekulasi dalam menyampaikan jawaban, atau karena adanya ketidakpedulian peserta Tes dalam mengerjakan soal.

Ditinjau dari Validitas Isi yang berfungsi untuk mengukur relevansi instrumen tes dalam mencapai tujuan yang ditetapkan, hasil analisis statistik menunjukkan 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. Berdasarkan parameter tersebut maka dapat disimpulkan bahwa 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.

Item fit berfungsi untuk mendeteksi kesalahan yang terjadi selama fase penentuan standar pengukuran yang akurat pada pengembangan instrumen tes. Melalui parameter daya pembeda dapat teridentifikasi baik atau tidaknya instrumen selanjutnya *item fit statistic* dapat mengidentifikasi masalah sehingga instrumen dapat diperbaiki sebelum diuji cobakan ke peserta didik. *Person fit* berfungsi untuk mengidentifikasi ada atau tidaknya penyimpangan pola respon dari peserta tes. Penyimpangan terjadi karena ketidakpedulian perilaku peserta dalam mengerjakan soal, seperti mencontek, kesalahan, kecerobohan, dan memiliki pemahaman yang berbeda tentang materi teks atau keberhasilan menebak jawaban.

Ditinjau dari Validitas isi yang berfungsi untuk mengukur relevansi instrumen tes dalam mencapai tujuan yang ditetapkan, hasil analisis statistik menunjukkan bahwa instrumen Tes pengembangan dapat mengukur hal yang ingin diukur. Dari hasil statistik outfit yang berada pada

rentang 0,756 hingga 1,434 atau $0,5 < MSQ < 1,5$. Implikasi kecocokan model terhadap pengukuran dapat dilihat berdasarkan MSQ. Jika $MSQ > 2,0$ bermakna merusak sistem pengukuran, $1,5 < MSQ < 2,0$ memiliki makna terhadap pengukuran: $0,5 < MSQ < 1,5$ berarti bermanfaat bagi pengukuran, dan $MNSQ < 0,5$ berarti tidak bermanfaat terhadap pengukuran meski tes tersebut tidak merusak sistem pengukuran. Sesuai parameter tersebut dapat disimpulkan bahwa 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.

Berdasar hasil analisis validasi konstruk aspek isi pada penelitian ini dari 224 peserta tes tampak bahwa seluruh butir secara umum dapat diterima sebagai butir yang baik kecuali butir nomor 13. Butir nomor 13 memiliki $p\text{-value} < 0,01$ dapat disimpulkan bahwa taraf signifikansi 0.01 butir nomor 13 tidak bisa diterima oleh model. Berdasar hasil analisis tingkat kesulitan dari rentang -8.612 hingga +3.172. Terdapat 22 butir soal kategori mudah, 1 butir soal tergolong sedang, 7 butir soal tergolong sukar. Butir soal dengan nilai kesukaran terendah terdapat pada soal nomor 12 sebesar -8.612 dan nilai kesukaran tertinggi pada nomor 28 yaitu mencapai +3.172. Berdasarkan perhitungan secara statistik dan ilustrasi Item Map dapat diketahui bahwa Instrumen Tes pengembangan yang disusun dalam penelitian ini memiliki item-item yang dapat digunakan untuk meningkatkan kemampuan literasi membaca peserta didik karena berada tingkat sedang dan tinggi.

Hasil uji person fit dari 224 respons tes literasi membaca peserta didik SMP ada 11 peserta tes yang mengalami respon yang menyimpang dari model. Hal ini terlihat dari ke 11 peserta tes tersebut tidak memenuhi sebanyak dua (p value dan outfit MSQ) dari ketiga kriteria person fit. Hasil validitas konstruk Aspek Substantif dari 11 peserta atau 4,91% dinyatakan menyimpang. Hal itu juga menunjukkan bahwa 95,08% peserta didik telah memenuhi Validasi Aspek substantif berdasar Model Rasch. Hasil analisis tersebut juga mengandung arti bahwa 95,08% peserta tes mengerjakan soal secara jujur dan tidak melakukan tindakan menyimpang yang dapat mengganggu kesimpulan, sehingga hasilnya dapat dipertanggungjawabkan. Sesuai grafik yang menghubungkan tingkat kemampuan peserta didik dengan nilai fungsi informasi, diketahui bahwa peserta didik memiliki kemampuan dalam menyelesaikan Tes dengan benar pada semua tingkat kesulitan tes. Hal tersebut mengidentifikasi bahwa instrumen Tes hasil pengembangan dapat digunakan.

Setelah dikaji dari validitas aspek struktural, hasil uji unidimensi menunjukkan bahwa butir-butir instrumen dalam penelitian ini memiliki nilai eigenvalue sebesar 0,5842. Nilai ini lebih besar dari 0,05, sehingga hipotesis nol (H_0) dapat diterima. Jika H_0 diterima, maka nilai eigenvalue yang kedua dan seterusnya akan lebih kecil daripada nilai eigenvalue yang pertama. Kondisi ini mengindikasikan bahwa tes hanya mengandung satu dimensi. Dengan demikian, dapat disimpulkan bahwa tes literasi membaca pada mata pelajaran Bahasa Inggris kelas VIII bersifat unidimensi.

Sesuai dengan hasil uji invariansi pengukuran uji Anderson LR test diketahui nilai p value sebesar 0,431 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.

Hasil nilai Person Separation reliability sebesar 0.2565 maka nilai separasi person untuk tes tersebut adalah 0.674. Dari nilai separasi person tersebut dapat diketahui bahwa klasifikasi peserta tes yang didapat hanya 1 (pembulatan dari 0.67). Artinya instrumen yang telah dibuat hanya membedakan peserta tes dalam dua kategori saja yaitu literat dan tidak literat. Konsekuensinya bahwa hasil tes ini hanya membedakan peserta tes menjadi dua kelompok yaitu peserta tes yang telah memiliki kecukupan minimal dalam mengerjakan soal Bahasa Inggris dan yang belum memiliki kecukupan minimal Bahasa Inggris.

Dari hasil penelitian diketahui ada satu butir soal yang tidak layak digunakan sebagai instrumen pengukuran kemampuan literasi membaca yaitu butir yang tidak cocok dengan model di butir soal nomor 11.

SIMPULAN

Berdasarkan hasil analisis menggunakan *Rasch Model* dapat disimpulkan bahwa terdapat kebutuhan pengembangan tes literasi membaca bagi Guru dan Peserta Didik di UPTD SPF SMP N 10 Tegal untuk meningkatkan kemampuan literasi membaca. Hasil analisis validitas konstruk tes

Mata Pelajaran bahasa Inggris berstandar PISA dalam penelitian ini diperoleh hasil bahwa 1 butir soal tidak terpakai dan 13 butir soal layak untuk digunakan, sehingga dapat dimanfaatkan untuk mengukur kemampuan literasi membaca peserta didik.

DAFTAR PUSTAKA

- Adawiyah, R., & Wisudawati, A. W. (2017). Pengembangan Instrumen Tes Berbasis Literasi Sains : Menilai Pemahaman Fenomena Ilmiah Mengenai Energi. *Indonesian Journal of Curriculum*, 5(2).
- Ahmadi, F., & Ibda, H. (2022). Media Literasi Sekolah. In *CV Pilar Nusantara*.
- Ardiyanti, D., & Dinni, S. M. (2018). Aplikasi Model Rasch dalam Pengembangan Instrumen Deteksi Dini Postpartum Depression. *Jurnal Psikologi*, 45(2). <https://doi.org/10.22146/jpsi.29818>
- Arikunto, S. (2002). Metodologi Penelitian Suatu Pendekatan Proposal. 2017.
- Arikunto, Suharsimi. (2021). Dasar-Dasar Evaluasi Pendidikan Edisi 3 - Google Books. In *Bumi Aksara* (p.334).https://www.google.co.id/books/edition/Dasar_Dasar_Evaluasi_Pendidikan_Edisi_3/j5EmEAAAQBAJ?hl=en&gbpv=&dq=suharsimi+arikunto+dasardasar+evaluasi+pendidikan&printsec=frontcover
- Bashoor, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2). <https://doi.org/0.283/pep.v22i2.9590>
- Berti, A. (2019). Peningkatan Keterampilan Membaca Cerpen Dengan Metode P2R dan Berpikir-Berpasangan-Berbagi. *Jurnal Pendidikan Tambusai*, 3(6).
- Dilla, R. F. (2019). Penilaian Aspek Pengetahuan Melalui Jenis Penialian Tes Di TK IL-Fadhilah Kabupaten Sleman DIY. *Jurnal Kajian Anak (J-Sanak)*, (0). <https://doi.org/0.2427/j-sanak.v10.3>
- Diputera, A. M. (2019). Teori penilaian tes esai atau uraian. *Journal Reseapedia*, ().
- Farahiba, A. S. (2022). Pengembangan Instrumen Tes Literasi Peserta Didik Pada Materi Teks Anekdote. *Jurnal Dimensi Pendidikan Dan Pembelajaran*, 0(2). <https://doi.org/0.24269/dpp.v0i2.4554>
- Fikriyah, F., Rohaeti, T., & Solihati, A. (2020). Peran Orang Tua dalam Meningkatkan Literasi Membaca Peserta Didik Sekolah Dasar. *Dwija Cendekia: Jurnal Riset Pedagogik*, 4(). <https://doi.org/0.2096/jdc.v4i.43937>
- Gall, M. D., & Borg, W. R. (1989). Educational Research an Introduction fourth edition. In *Longman Inc.*
- 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 Ketidakwaajaran 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)
- 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*