

Pemodelan Rasch Bidang Sosial rev2- press copy.pdf

WORD COUNT

16254

TIME SUBMITTED

27-JUL-2024 04:32AM

PAPER ID

110701171

- Dr. Moh. Irma Sukarelawan, M.Pd.
- Toni Kus Indratno, M.Pd.Si.

Validasi Instrumen Penelitian Menggunakan **PEMODELAN RASCH** di Bidang Sosial



Validasi instrumen penelitian menggunakan

PEMODELAN RASCH

di bidang sosial

16

Dr. Moh. Irma Sukarelawan, M.Pd. | Toni Kus Indratno, M.Pd.Si.



Validasi Instrumen Penelitian Menggunakan Pemodelan Rasch di Bidang Sosial
© 2024 by Universitas Ahmad Dahlan. All right reserved.

Penyusun:

Dr. Moh. Irma Sukarelawan, M.Pd
Toni Kus Indratno, M.Pd.Si.

19

Hak cipta dilindungi oleh undang-undang, jika hendak memperbanyak atau memperbaiki isi buku ini hendaknya memperoleh ijin terlebih dahulu, baik dari penulis maupun dari penerbit.

Desain Sampul : Pustaka Pranala
Tata Letak : Toni Kus Indratno

Cetakan Pertama, Mei 2024
x + 96 hlm.; 182 x 257 mm.

ISBN : **978-623-5427-39-3**
Buku Monograf

38 erbit Pustaka Pranala

Gg. Nogobondo III No.500a, Rejowinangun, Kec. Kotagede, Kota Yogyakarta, Daerah Istimewa Yogyakarta 55171

Pengantar

Puji syukur kami panjatkan ke hadirat Allah Tuhan Yang Maha Tinggi, yang telah memberikan rahmat dan karunia-Nya sehingga buku berjudul "Validasi Instrumen Penelitian Menggunakan Pemodelan Rasch di Bidang Sosial" ini dapat terselesaikan dengan baik. Buku ini hadir sebagai jawaban atas kebutuhan akan metode validasi instrumen yang andal dalam penelitian sosial, khususnya dalam konteks pendidikan di Indonesia.

Dalam proses pendidikan, terutama di tingkat menengah dan tinggi, evaluasi yang akurat terhadap kemampuan siswa sangatlah penting. Instrumen yang valid dan reliabel merupakan kunci untuk mendapatkan data yang sah, yang pada akhirnya akan mendukung pengambilan keputusan yang tepat dalam bidang pendidikan. Oleh karena itu, buku ini menyajikan pendekatan pemodelan Rasch sebagai metode yang unggul dalam validasi instrumen penelitian.

Pemodelan Rasch bukanlah metode yang baru, namun penerapannya dalam konteks pendidikan di Indonesia masih memerlukan banyak perhatian dan adaptasi. Buku ini memberikan penjelasan yang komprehensif tentang bagaimana pemodelan Rasch dapat digunakan untuk validasi berbagai instrumen penelitian, mulai dari instrumen kesadaran metakognitif hingga persepsi belajar *online*. Kami berharap buku ini dapat menjadi referensi bagi para peneliti, akademisi, dan praktisi pendidikan yang ingin meningkatkan kualitas penelitian mereka melalui validasi instrumen yang lebih baik.

57

Buku ini terdiri dari beberapa bab yang masing-masing membahas validasi instrumen yang berbeda, yaitu Jr.MAI untuk kesadaran metakognitif, POSTOL untuk persepsi belajar online, HELAM untuk penilaian e-learning, dan I-PMQ untuk motivasi belajar fisika. Setiap bab dilengkapi dengan metodologi yang detail, hasil dan pembahasan yang mendalam, serta kesimpulan yang jelas, sehingga pembaca dapat memahami proses validasi instrumen dengan menggunakan pemodelan Rasch secara menyeluruh.

8

Kami menyadari bahwa buku ini masih jauh dari sempurna. Oleh karena itu, kami sangat terbuka terhadap kritik dan saran dari para pembaca untuk perbaikan di masa mendatang. Harapan kami, buku ini dapat memberikan manfaat bagi perkembangan penelitian sosial dan pendidikan.

8

Akhir kata, kami mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan dalam penyusunan buku ini. Semoga buku ini dapat memberikan kontribusi penelitian-penelitian yang akan datang.

Selamat membaca!

Daftar Isi

Bab 1 Validasi Instrumen Kesadaran Metakognitif Jr.MAI	1
A. Perumusan Masalah	1
B. Metodologi.....	4
C. Hasil dan Pembahasan	5
D. Kesimpulan	18
E. Potensi penelitian mendatang.....	19
F. Daftar Pustaka	19
Bab 2 Validasi Instrumen Persepsi Belajar Online POSTOL	25
A. Perumusan Masalah	25
B. Metodologi.....	30
C. Hasil dan Pembahasan	32
D. Kesimpulan	38
E. Potensi penelitian mendatang.....	39
F. Daftar Pustaka	39
Bab 3 Validasi Instrumen Model Penilaian E-Learning HELAM	45
A. Perumusan Masalah	45
B. Metodologi.....	48
C. Hasil dan Pembahasan	51
D. Kesimpulan	67
E. Potensi penelitian mendatang.....	68
F. Daftar Pustaka	68
Bab 4 Validasi Instrumen Motivasi Belajar Fisika I-PMQ	75
A. Perumusan Masalah	75
B. Metodologi.....	77
C. Hasil dan Pembahasan	78

D. Kesimpulan	85
E. Potensi penelitian mendatang.....	86
F. Daftar Pustaka	87
Lampiran.....	91

Daftar Tabel

Tabel 1. Ringkasan fungsi skala rating dalam Jr.MAI.....	6
Tabel 2. Ringkasan statistik berdasarkan parameter Rasch	8
Tabel 3. Komparasi MNSQ antara mahasiswa Singapura dengan Indonesia.....	11
Tabel 4. Statistik demografi	31
Tabel 5. Ringkasan statistik POSTOL	32
Tabel 6. Karakteristik item di POSTOL	33
Tabel 7. Statistik fungsionalitas skala peringkat Likert lima poin.....	35
Tabel 8. Karakteristik responden	49
Tabel 9. Fungsi Skala Rating pada HELAM	51
Tabel 10. Ringkasan Statistik HELAM dalam konteks Indonesia.....	53
Tabel 11. <i>Item fit</i> HELAM dalam konteks Indonesia	56
Tabel 12. Model kecocokan PMQ 14-item.....	83

Daftar Gambar

Gambar 1. Kurva probabilitas untuk skala Likert 5 poin dalam Jr.MAI	7
Gambar 2. Probabilitas item lintas gender	15
Gambar 3. DIF size berdasarkan gender	15
Gambar 4. Person-item map of the Jr.MAI	17
Gambar 5. Peta orang-item berdasarkan <i>Andrich threshold</i>	18
Gambar 6. Proses adaptasi lintas budaya	28
Gambar 7. Kurva probabilitas skala peringkat Likert 5 poin POSTOL	34
Gambar 8. Kurva Probabilitas skala Likert HELAM	52
Gambar 9. Wright Map dari HELAM.....	54
Gambar 10. Bagan Buble-fit item berdasarkan statistik Fit	59
Gambar 11. DIF Plot Item di HELAM Berdasarkan Gender	62
Gambar 12. Ringkasan statistik PMQ	79
Gambar 13. Peta Wright motivasi belajar siswa.....	81
Gambar 14. Kurva probabilitas fungsionalitas skala peringkat Likert di PMQ ...	85

Daftar Lampiran

Lampiran 1. Instrumen Jr.MAI.....	91
Lampiran 2. Instrumen POSTOL.....	93
Lampiran 3. Instrumen I-PMQ.....	95

BAB 1

Validasi Instrumen Kesadaran Metakognitif Jr.MAI

A. Perumusan Masalah

Saat ini, berbagai literatur telah melaporkan peran kritis metakognisi dalam proses belajar siswa. Kemampuan memantau dan mengontrol pembelajaran berkorelasi positif dengan keberhasilan belajar, peningkatan prestasi akademik, serta kesehatan dan kesejahteraan siswa [1]–[3]. Misalnya, strategi metakognitif berkorelasi positif dengan nilai tes siswa [4], [5]. Selain itu, penggunaan praktis keterampilan dan strategi metakognitif telah meningkatkan pembelajaran di berbagai tingkat pendidikan [1], [6]–[8].

Taksonomi metakognitif telah berkembang dalam empat dekade terakhir. Flavell menjadi inisiator dalam memperkenalkan konsep metakognitif. Pada awal kemunculannya, metakognitif dipahami sebagai “pemikiran tentang fenomena kognitif” [9]. Dengan kata lain, metakognitif dapat dipandang sebagai kesadaran atau aktivitas kognitif seseorang tentang proses berpikir atau segala sesuatu yang berhubungan dengannya [10]. Flavell [9] mengusulkan struktur metakognitif yang terdiri dari 4 komponen utama (*metacognitive knowledge, metacognitive experiences, goal, dan action*). Sekitar satu setengah dekade kemudian, Schraw dan

Dennison mengusulkan struktur metakognitif yang terdiri dari 2 komponen utama: Pengetahuan Metakognitif (MK) dan Pengalaman Metakognitif (ME). Komponen MK terdiri dari tiga sub-komponen, yaitu: *declarative knowledge*, *procedural knowledge*, dan *conditional knowledge*. Pada saat yang sama, komponen ME terdiri dari 5 sub-komponen, yaitu: *planning*, *monitoring*, *information management*, *debugging*, dan *evaluation* [11]. Usulan ini merupakan penyempurnaan dari struktur faktor yang dikemukakan oleh Brown [12].

Dalam kurikulum pendidikan nasional Indonesia, siswa pada jenjang SMA harus memiliki kemampuan metakognitif [13]. Oleh karena itu, diperlukan instrumen standar untuk memudahkan tugas guru dalam menilai kemampuan tersebut secara akurat. Beberapa kuesioner metakognitif telah dikembangkan dan diterapkan di lapangan [14]. Hal ini sebagai hasil dari adanya beberapa taksonomi metakognitif yang telah diusulkan sebelumnya [9], [11], [15]. Kesadaran metakognitif sifatnya sulit untuk diamati dan dinilai. Kuesioner *self-report* adalah cara yang paling efektif dan efisien dan paling tidak bermasalah untuk mengevaluasi kesadaran metakognitif seseorang [2], [16].

Komponen yang diusulkan oleh Schraw dan Dennison telah menghasilkan banyak kuesioner *self-report* untuk menilai kesadaran metakognitif, misalnya *Metacognitive Awareness Inventory (MAI)* [11], *Junior Metacognitive Awareness Inventory (Jr.MAI)* Versi A dan B [17], dan *Physics Metacognition Inventory (PMI)* [18], [19].

Sperling et al. [17] telah mengembangkan Jr.MAI versi B. Berdasarkan *exploratory factor analysis*, 18 item Jr.MAI memiliki *concurrent validity* dan membentuk 2 faktor metakognitif (Pengetahuan Kognisi dan Regulasi Kognisi). Laporan penelitian menunjukkan bahwa model 18 item cocok dengan dua faktor (lihat Tabel 3). Reliabilitas internal

Jr.MAI memiliki koefisien korelasi sebesar 0,82. Oleh karena itu, item-item di Jr.MAI dapat diandalkan. Temuan signifikan ini menarik minat para peneliti, termasuk kami, untuk mengkaji ulang Jr.MAI dalam konteks bahasa yang berbeda.

Tes pendidikan dan psikologis perlu ada versi *Multilanguage* karena minat tentang prestasi ilmiah dalam studi komparasi internasional dan psikologis lintas budaya meningkat [16]. Mengingat bahwa instrumen *self-report* Jr.MAI dimaksudkan untuk mengukur metakognisi siswa di Amerika Serikat, aplikasi dalam konteks yang berbeda memerlukan perhatian khusus. Beberapa peneliti telah mengkonfirmasi penggunaan Jr.MAI di pada konteks Bahasa yang berbeda, misalnya versi Turki [16], Korea [20], dan Singapore [3], [21].

Penggunaan instrumen kesadaran metakognitif di Indonesia telah banyak dilaporkan [13], [22]–[26]. Namun, ada keterbatasan literatur yang melaporkan tentang proses adaptasi dan studi komprehensif sifat psikometrik instrumen, terutama Jr. MAI versi B. Oleh karena itu, studi tentang sifat psikometrik Jr. MAI versi B perlu dilaporkan. Laporan ini akan memastikan kesesuaian dan keakuratan informasi yang dikumpulkan bila digunakan dalam konteks Indonesia.

46

Teknik analisis faktor (*exploratory factor analysis* dan *confirmatory factor analysis*) telah digunakan untuk menetapkan validitas konstruk Jr.MAI versi B [16], [20], [21]. Pengujian fungsi *rating scale*, unidimensionalitas, analisis bias melalui pengujian *Differential Item Functioning* (DIF), kualitas item (kesulitan item dan kemampuan responden) merupakan keterbatasan yang tidak dilaporkan pada penelitian sebelumnya. Oleh karena itu, diperlukan teknik analisis Rasch yang didasarkan pada *Item Response Theory* untuk mengisi gap ini. Sejauh pengamatan kami, didukung oleh Craig et al. [2], ada sedikit laporan

tentang analisis Rasch untuk mengevaluasi sifat psikometrik penggunaan Jr.MAI di Indonesia. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi sifat psikometri Jr. MAI dalam konteks Bahasa Indonesia dengan menggunakan teknik analisis Rasch. Dengan demikian, diharapkan para guru atau konselor dapat menggunakan Jr. MAI dalam mengukur metakognisi siswa di Indonesia.

B. Metodologi

Jr.MAI versi Indonesia diberikan kepada beberapa sekolah menengah atas di kota Yogyakarta, Indonesia. Kami memilih Yogyakarta karena merupakan pusat pendidikan dan identik sebagai “kota pelajar”. Banyak siswa yang datang dari berbagai daerah di Indonesia untuk belajar di kota ini. Sehingga peneliti berasumsi ada heterogenitas siswa di Yogyakarta. Peneliti meminta izin kepada kepala sekolah dan guru. Persetujuan etik juga diberikan dari Universitas Negeri Yogyakarta. Dengan bimbingan dan pengawasan peneliti dan guru, 351 siswa telah berpartisipasi dan mengisi kuesioner *online* menggunakan *stratified random sampling*. Kami menjalankan *screening* data untuk mengecualikan *outlier* sebelum analisis data, 55 *outlier* dikeluarkan dari dataset. Oleh karena itu, 296 siswa (Laki-laki = 45,9% dan Wanita = 54,1%) dianalisis dengan pengukuran Rasch menggunakan *software Winsteps*. Jumlah peserta dalam analisis ini cukup untuk stabilitas data untuk ukuran sampel, lebih dari 250 responden [27], [28].

Instrumen Jr.MAI [17] telah diadaptasi dan diterjemahkan ke dalam versi bahasa Indonesia oleh penerjemah bahasa Indonesia menggunakan *back-forward translation*. Konteks lintas budaya telah dinilai dalam mengadaptasi kuesioner pada konteks Indonesia [29]. Kuesioner Jr.MAI terdiri dari 18 item yang dibagi menjadi dua konstruk: *knowledge of*

cognition (KoC) dan *regulation of cognition* (RoC). Pada Jr.MAI, KoC dan RoC masing-masing terdiri dari 9 item (lihat Tabel 3). Setiap item menggunakan lima kategori skala Likert mulai dari 1 (tidak pernah) sampai 5 (selalu). Winsteps versi 4.6.1 digunakan untuk menganalisis dataset berdasarkan pengukuran Rasch. Kami lebih suka menggunakan pengukuran Rasch karena pengukuran Rasch dapat mengatasi beberapa keterbatasan analisis Classical Test Theory (CTT) seperti data yang hilang dalam analisis, parameter reliabilitas hanya menggunakan alpha Cronbach, ketergantungan item dan person yang mungkin tidak dapat diandalkan dan valid dalam konteks penelitian lain [30].

Analisis data dimulai dengan *screening* data tanggapan peserta. Kami menerapkan analisis *rating scale* menggunakan pemodelan Rasch untuk melakukan analisis data. Fungsi *rating scale* ditinjau dari peningkatan *average observation*, *Andrich Threshold values*, dan *probability curves* [31], [32]. Reliabilitas instrumen ditentukan berdasarkan koefisien *alpha Cronbach*, parameter reliabilitas person dan item. Item fit ditentukan dari nilai statistik *Infit and Outfit MNSQ*, *wright map*, *Local independence*, dan *unidimensionality*. Bias item Jr.MAI berdasarkan *gender* ditentukan berdasarkan *Differential Item Functioning* (DIF) [33].

C. Hasil dan Pembahasan

Screening Rating Scale

Sifat psikometrik dari 18 item Jr.MAI dianalisis menggunakan software WINSTEPS 4.6.1. Analisis data dimulai dengan *screening person* dan *rating scale*. Dalam proses *screening* data, 55 *outlier* telah terdeteksi. *Outliers* adalah siswa yang jawabannya mencurigakan dan tidak sesuai. Nilai MNSQ person berada di luar kriteria yang dapat diterima (0,5 hingga

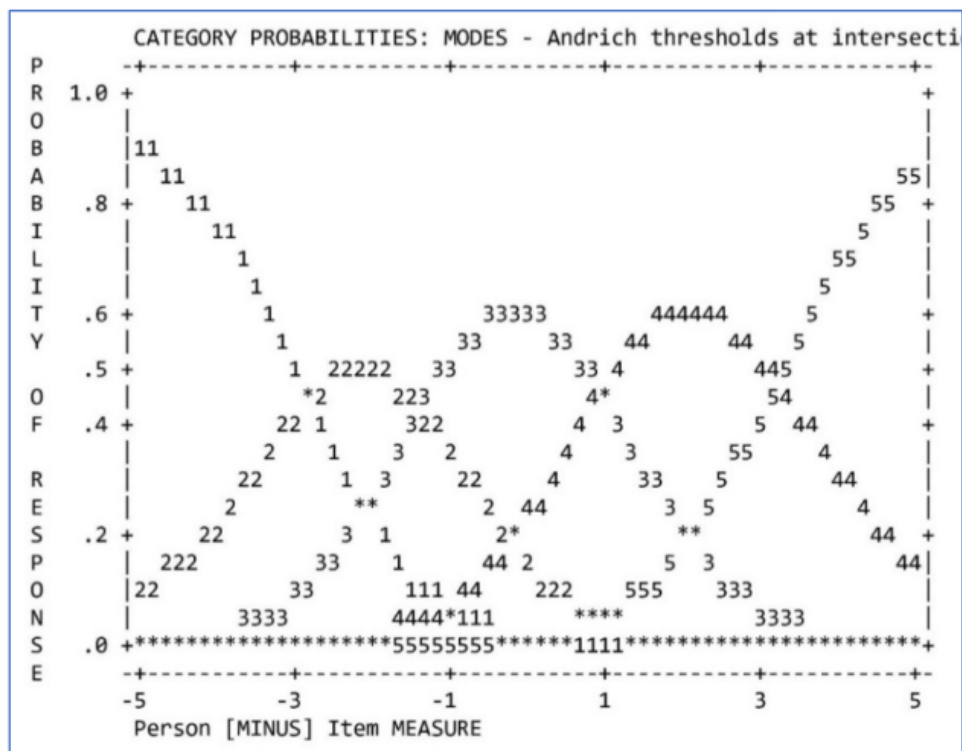
1,5) yang diindikasikan sebagai orang yang tidak sesuai atau *outlier* (Andrich, 2018; Bond et al. 2015). Setelah *screening person*, kami mengevaluasi *rating scale* yang digunakan di Jr.MAI. Analisis fungsi pilihan pada *rating scale* yang disediakan sangat penting sebagai elemen kualitas psikometrik skala [32]. *Rating scale* yang baik adalah jika pilihan yang diberikan tidak membingungkan responden. Tabel 1 menunjukkan ringkasan parameter yang digunakan untuk menilai fungsi pilihan pada *rating scale* di Jr.MAI.

Tabel 1. Ringkasan fungsi skala rating dalam Jr.MAI

43	Kategori	Counts	Observed Average	Andrich Threshold
	1 (tidak pernah)	37	-0.47	-
	2 (Jarang)	265	-0.30	-2.77
	3 (Terkadang)	1450	0.93	-1.38
	4 (Sering)	2439	1.90	0.94
	5 (Selalu)	1137	3.00	3.22

Berdasarkan Tabel 1, tampak bahwa *observed counts* memiliki distribusi unimodal. *Observed Average* meningkat secara monoton dari -0,47 sampai +3,00 logit. Indikator lain yang perlu diperhatikan untuk melihat berfungsinya pilihan skala adalah *Andrich Threshold* [32], [35]. Nilai *Andrich Threshold* meningkat secara monoton dari *NONE* menjadi 3,22. Pilihan skala yang baik jika setiap level telah meningkat lebih dari satu logit dalam parameter *Andrich Threshold* [3]. Terdapat peningkatan pada setiap *rating scale* yang diberikan minimal 1,39. Selain itu, pemeriksaan fungsi skala dapat melalui kurva probabilitas (Gambar 1). Semua Kategori pada kurva probabilitas memiliki puncaknya masing-masing di sepanjang sumbu ukur. Hal ini menunjukkan kesesuaian dengan pola yang direkomendasikan [36]. Temuan ini sedikit berbeda dengan penggunaan Jr.MAI pada siswa Singapura. Penggunaan skala Likert “Rarely” dalam studi Ning (2018) memiliki puncak di bawah kurva

probabilitas untuk kategori “sometimes” dan “never”. Jadi skala penilaian Jr.MAI yang digunakan untuk siswa Singapura perlu disederhanakan menjadi skala Likert 4 poin. Berdasarkan nilai *Observed Average*, *Andrich Threshold*, dan *kurva probabilitas*, dapat dinyatakan bahwa skala Likert 5 poin yang digunakan di Jr.MAI untuk siswa Indonesia dapat berfungsi dengan baik.



Gambar 1. Kurva probabilitas untuk skala Likert 5 poin dalam Jr.MAI

Instrument Reliability

Setelah *screening* person dan *rating*, kami menghitung statistik untuk kuesioner Jr.MAI berdasarkan parameter item dan person dalam pemodelan Rasch. Tabel 2 mewakili ringkasan statistik kuesioner Jr.MAI untuk item dan person berdasarkan parameter Rasch.

Tabel 2 menunjukkan bahwa rata-rata kemampuan person adalah 1,74 logit, di atas rata-rata (0 logit), dan kesulitan item berada pada kisaran rata-rata (0 logit). Pemisahan item menunjukkan bahwa kuesioner Jr.MAI memiliki 13 tingkat kesulitan item yang berbeda. Pemisahan person membuktikan bahwa setidaknya ada dua level person dalam penelitian ini, yaitu siswa berkemampuan tinggi dan rendah. Secara keseluruhan, data memiliki Chi-kuadrat (χ^2) = 10072,06 (df= 10090), $p > 0,05$ menunjukkan distribusi normal telah tercapai.

Tabel 2. Ringkasan statistik berdasarkan parameter Rasch

	Persons	Item
N	296	18
Mean	68.7	1131
Measure	1.74	0
SD	0.87	0.92
SE	0.06	0.22
Mean Outfit ZSTD	-0.07	-0.10
Mean Outfit MNSQ	1.00	1.00
Separation	2.23	9.68
Strata	3.31	13.24
Reliability	0.83	0.99
Cronbach's Alpha	0.85	
Chi-squared (χ^2)	10072.06 (df= 10090)	
Probability	0.5484 *	

*Normally distributed

Reliabilitas instrumen diestimasi berdasarkan item dan person (lihat Tabel 2). Rata-rata *output person* sebesar 1,74 logit. Hal ini menunjukkan kecenderungan responden untuk setuju pada berbagai atribut di Jr.MAI. Data item dan person digunakan untuk melihat kesesuaian penggunaan item pada Jr.MAI dan kesesuaian statistik responden. Interaksi person-item pada penggunaan Jr.MAI sudah tepat dan reliabel karena memiliki nilai Cronbach alpha sebesar 0,85. Konsistensi responden baik, dan kualitas item di Jr.MAI istimewa [37]. Hal ini didukung oleh nilai reliabilitas person dan item masing-masing 0,83 dan 0,99. Ketiga nilai reliabilitas

tersebut menunjukkan item-item dalam Jr.MAI dapat mendefinisikan variabel laten dengan baik [38]. Nilai pemisahan person dan item, yang diwakili dalam strata, masing-masing adalah 3,31 dan 13,24. Nilai ini menunjukkan bahwa Jr.MAI memiliki kemampuan yang baik untuk mengklasifikasikan person dan item.

Unidimensionality and local independence

Unidimensionalitas skala Jr.MAI telah ditentukan dengan menggunakan *residual Principal Component Analysis*. Unidimensionalitas menjelaskan bahwa instrumen tersebut bersifat unidimensi dalam mengukur faktor laten dalam penelitian ini, *Metacognitive Awareness*. Jr.MAI dapat mencapai kriteria validitas dalam mengukur faktor laten atau unidimensi jika *score of raw variance explained by measure* lebih dari 30% [39]. Nilai *raw variance explained by measures* kuesioner Jr.MAI adalah 42,8%. Nilai ini membuktikan adanya unidimensionalitas yang baik pada skala Jr.MAI yang mengukur satu dimensi. Temuan ini juga mendukung dan memperkuat unidimensionalitas Jr.MAI pada siswa Singapura [3]. *Local independence* menjelaskan bahwa setiap item dalam kuesioner Jr.MAI tidak bergantung. Instrumen dapat mencapai kriteria *local independence* jika korelasi antar item lebih kecil dari 0,3. *Raw residual correlation* antara item kuesioner Jr.MAI di bawah 0,3, yang membuktikan bahwa kuesioner bebas dari masalah *local dependence*.

Item Fit

Nilai statistik *infit* dan *outfit MNSQ* pada Tabel 3 digunakan untuk mengukur kesesuaian individual item di Jr.MAI [40]. Item yang cocok akan memberikan kontribusi yang baik dalam mendefinisikan sebuah *common construct* [36]. Dalam pemodelan Rasch, nilai *infit* dan *outfit MNSQ* yang

ideal adalah 1. Nilai 0,5 - 1,5 adalah kisaran penerimaan yang wajar yang menunjukkan nilai produktif untuk pengukuran (Andrich & Marais, 2019; Bond & Fox, 2015; Wright & Linacre, 1994). Semua, 18 item Jr.MAI memiliki nilai *infit*, dan *outfit* MNSQ berada dalam kisaran penerimaan. Ini menunjukkan kecocokan dalam pola respon terhadap item target dan kecocokan antara kemampuan person dan kesulitan item. Berbeda dengan temuan yang dilansir Ning. Dua item (KoC6 dan RoC1 memiliki sifat psikometrik yang tidak mencukupi (lihat Tabel 3) [3]. Selain itu, nilai *PT-Measure Corr* skala Jr.MAI bergerak ke arah positif dari 0,41 sampai 0,67, seperti yang ditunjukkan pada Tabel 3. Hal ini menunjukkan kesesuaian semua item terhadap variabel laten yang disepakati [38], sehingga dapat disimpulkan bahwa 18 item Jr.MAI dapat diterapkan untuk mengukur kesadaran metakognitif siswa SMA di Indonesia.

Tabel 3. Komparasi MNSQ antara mahasiswa Singapura dengan Indonesia.

Item	Statemen	Mahasiswa Indonesia				Mahasiswa Singapura				PT-Mea. Corr.	Measure
		Infit		Outfit		Infit		Outfit			
		MNSQ	MNSQ	MNSQ	MNSQ	MNSQ	MNSQ	MNSQ	MNSQ		
KoC1	Saya tahu ketika saya memahami sesuatu.	0.85	0.85	0.85	0.85	0.96	0.92	0.96	0.92	0.57	0.82
KoC2	Ketika saya menyukai suatu topik, saya mempelajarinya lebih mendalam	0.95	0.94	0.94	0.94	1.18	1.18	1.18	1.18	0.58	-0.75
KoC3	Informasi-informasi penting dalam suatu topik sangat saya perhatikan.	0.79	0.79	0.79	0.79	0.90	0.93	0.90	0.93	0.60	0.15
KoC4	Saya dapat mengkondisikan diri saya untuk belajar ketika saya butuh.	1.04	1.07	1.07	1.07	0.80	0.86	0.80	0.86	0.54	-0.11
KoC5	Saya bisa belajar maksimal ketika sudah mengetahui sesuatu dari topik itu	1.23	1.23	1.23	1.23	1.17	1.18	1.17	1.18	0.53	-0.41
KoC6	Saya mengerti apa yang diharapkan guru pada saya untuk dipelajari	1.28	1.30	1.30	1.30	1.41	1.55	1.41	1.55	0.43	2.22
KoC7	Cara belajar yang berhasil saya gunakan sebelumnya, akan saya gunakan kembali	1.12	1.12	1.12	1.12	0.85	0.88	0.85	0.88	0.44	-0.25
KoC8	Saya menggunakan kekuatan-kekuatan belajar yang saya miliki untuk menutupi kelemahan-kelemahan.	0.81	0.80	0.80	0.80	0.78	0.80	0.78	0.80	0.67	-0.53

Item	Statemen	Mahasiswa				PT-Mea. Corr.	Measure
		Indonesia		Singapura			
		Infit MNSQ	Outfit MNSQ	Infit MNSQ	Outfit MNSQ		
KoC9	Tanpa melalui proses berpikir, saya kadang-kadang langsung menggunakan strategi-strategi belajar.	0.97	0.97	1.17	1.33	0.50	-0.31
RoC1	Setelah menyelesaikan tugas sekolah, saya bertanya kepada diri sendiri apakah saya sudah mempelajari apa yang sebenarnya mau saya pelajari	1.10	1.11	1.53	1.64	0.50	-0.10
RoC2	Saya memikirkan apa yang sebenarnya ingin saya pelajari dari satu topik itu sebelum saya mulai mempelajarinya.	0.94	0.95	0.97	1.05	0.57	-1.29
RoC3	Ketika saya sedang mempelajari materi yang baru, saya bertanya pada diri sendiri seberapa baik saya melakukannya.	0.74	0.75	0.87	0.87	0.57	-0.19
RoC4	Ketika menyelesaikan tugas sekolah, saya mempertimbangkan beberapa cara penyelesaian lalu memilih cara terbaik	0.88	0.89	0.84	0.87	0.46	-0.11
RoC5	Setelah saya menyelesaikan tugas, saya bertanya pada diri sendiri apakah ada cara lain yang lebih mudah.	1.03	1.02	0.86	0.88	0.44	-0.79
RoC6	Untuk membantu memahami materi ketika belajar, saya membuat gambar-gambar atau diagram-diagram.	0.82	0.83	0.86	0.87	0.49	0.80

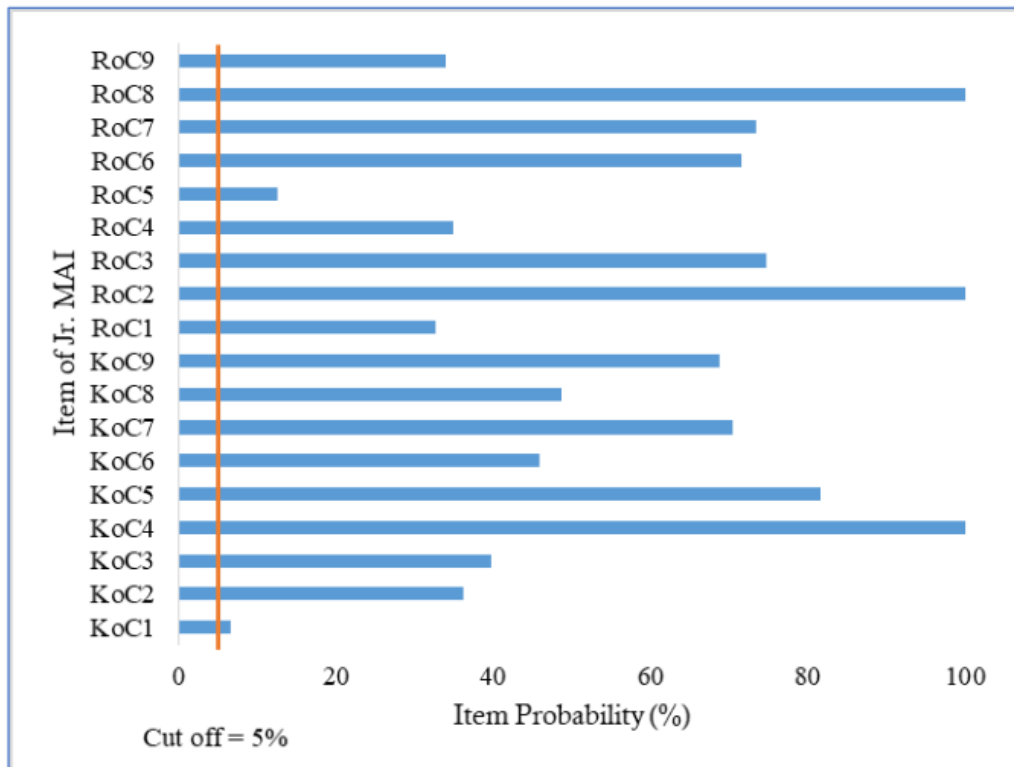
Item	Statemen	Mahasiswa Indonesia		Mahasiswa Singapura		PT-Mea. Corr.	Measure
		Infit	Outfit	Infit	Outfit		
		MNSQ	MNSQ	MNSQ	MNSQ		
RoC7	Strategi-strategi belajar yang saya gunakan bisa saja berbeda tergantung pada tugasnya.	0.83	0.82	0.98	1.01	0.54	-1.10
RoC8	Sebelum mulai mengerjakan tugas, saya memastikan apa saja yang harus dilakukan	1.24	1.22	1.14	1.14	0.54	-0.14
RoC9	Saya memeriksa kembali pekerjaan saya untuk memastikan bisa selesai tepat waktu	1.34	1.35	0.97	0.97	0.41	2.08

Differential Item Functioning (DIF)

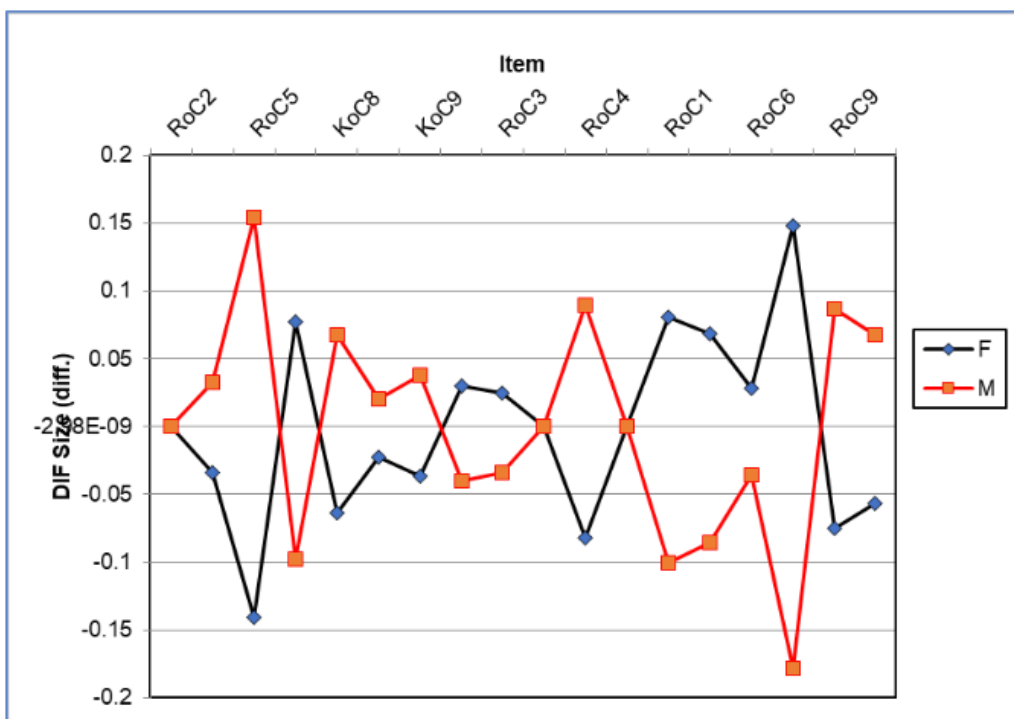
Salah satu ciri instrumen yang baik adalah tidak memiliki bias terhadap atribut tertentu dari responden. Analisis DIF dilakukan untuk melihat *trend item* di Jr.MAI pada atribut gender. Item memiliki bias gender jika nilai probabilitasnya kurang dari 5% (Sumintono & Widhiarso, 2014). ⁶² Seperti yang ditunjukkan pada Gambar 2, tidak ada nilai probabilitas < 5% sebagai indikasi bias butir terhadap gender. Kami juga menjalankan analisis DIF berdasarkan ukuran DIF (lihat Gambar 3). Jr.MAI membuktikan bahwa tidak ada bias gender pada setiap item karena tidak ada ukuran DIF yang memiliki skor $|DIF| \geq 0,43$ (sedikit hingga sedang) atau $|DIF| \geq 0,64$ logits (sedang hingga besar) [43]. Hasil ini menunjukkan item dalam Jr.MAI setara terhadap gender pria dan wanita, yang mendukung temuan lain dari Papini et al. [44]. Temuan ini juga sejalan dengan Jr.MAI pada mahasiswa Singapura [3].

Item KoC1 mendekati kriteria *cut-off* untuk nilai probabilitas ($p < 5\%$). Namun, kita dapat mengasumsikan bahwa KoC1 masih layak dipertahankan dalam kuesioner Jr.MAI. Secara keseluruhan, Gambar 2 menunjukkan bahwa tidak ada masalah bias di semua item.

Kami juga melakukan analisis DIF berdasarkan gender (Wanita (F)) dan Pria (M)). Gambar 3 mencoba mengilustrasikan DIF dalam Domain RoC. Namun, Gambar 3 menunjukkan tidak ada ukuran DIF yang bias secara substansial dalam item dalam kuesioner Jr.MAI yang terdeteksi, yang kurang dari 0,43 (kategori ringan hingga sedang).



Gambar 2. Probabilitas item lintas gender



Gambar 3. DIF size berdasarkan gender

Item and Person Distribution

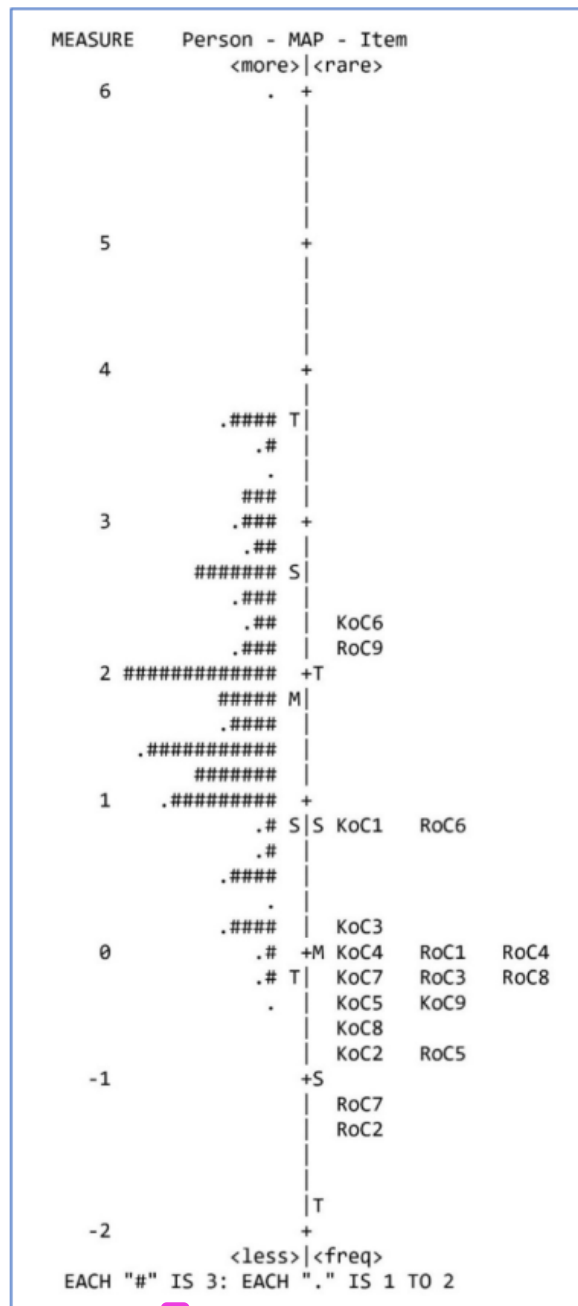
Hubungan antara person dan item divisualisasikan melalui peta person-item (*Wright map*). Model Rasch menampilkan kemampuan individu dan kesulitan item pada skala linier dalam satu kerangka acuan sehingga kemampuan individu dan kesulitan item tidak saling bergantung [46], [47]. Gambar 4 menunjukkan keadaan person dan item pada skala logit yang sama. Ini digunakan untuk membandingkan tingkat kesulitan item terhadap kemampuan individu tersebut.

Peta person-item pada Gambar 4 dibagi menjadi empat area. Area kanan-atas menunjukkan posisi item dengan tingkat kesulitan yang tinggi, atau siswa cenderung lebih sulit menyepakati pernyataan tersebut. Sedangkan area kanan-bawah menunjukkan item dengan tingkat kesulitan rendah, atau siswa cenderung mudah setuju dengan pernyataan yang ada. Area kiri-atas menunjukkan posisi person dengan tingkat kesadaran metakognitif tinggi, dan area kiri-bawah menunjukkan posisi person dengan tingkat metakognitif rendah.

Delapan belas item didistribusikan di sisi kanan peta. Item KoC6: “*Saya mengerti apa yang diharapkan guru pada saya untuk dipelajari.*” adalah yang paling sulit disepakati oleh siswa, dan item RoC2: “*Saya memikirkan apa yang sebenarnya ingin saya pelajari dari satu topik itu sebelum saya mulai mempelajarinya.*” yang paling mudah disetujui oleh siswa. Ada kesenjangan besar antara item KoC3 dan RoC9. Dengan demikian, untuk meningkatkan sensitivitas dan reliabilitas Jr.MAI, beberapa item perlu ditambahkan [48].

Skor rata-rata person lebih tinggi dari item. Hal ini menunjukkan bahwa rata-rata peluang siswa untuk memiliki kesadaran metakognitif lebih tinggi daripada rata-rata tingkat kesulitan item. Item dengan logit yang setara dengan logit person memiliki probabilitas 50% disetujui oleh siswa. Item yang berada di bawah logit person memiliki probabilitas lebih dari 50% untuk

disepakati siswa. Pada saat yang sama, item dengan logit di atas person memiliki probabilitas kurang dari 50% untuk disetujui oleh siswa [49].

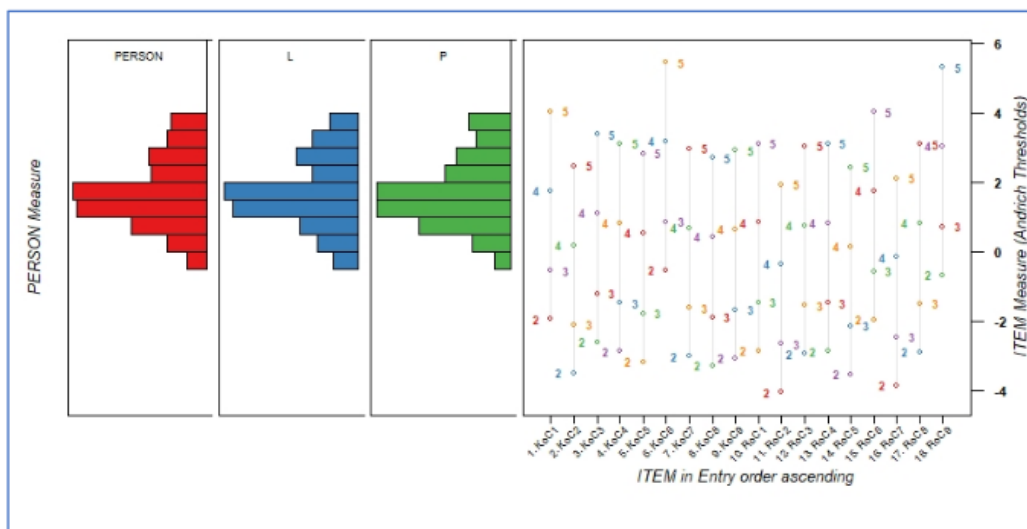


Gambar 4. Person-item map of the Jr.MAI

Peta Person-item menunjukkan dua item (RoC7 dan RoC2) yang paling mudah disetujui oleh siswa, tetapi RoC7 dan RoC2 masih di bawah dua unit logit standar deviasi, sehingga kita dapat mengasumsikan bahwa kedua item

ini bukan item yang tidak sesuai. Hasil ini menunjukkan bahwa regulasi kognisi siswa khususnya RoC7 dan RoC2 lebih tinggi dibandingkan domain item lainnya. Dari peta person-item Jr.MAI, ada ruang untuk perbaikan item dengan menambahkan item yang lebih sulit dijawab untuk mencakup semua kemampuan orang.

Untuk memastikan Jr.MAI dalam konteks Indonesia sesuai dengan penggunaan *rating scales*. Kami menjalankan peta person-item untuk person dan kelompok gender berdasarkan Andrich Threshold pada Gambar 5. Gambar 5 memberi kita pemahaman lebih lanjut tentang skor distribusi dalam kuesioner di mana semua skala penilaian bekerja dengan baik untuk semua person dan gender dalam kecocokan dan distribusi data.



Gambar 5. Peta orang-item berdasarkan *Andrich threshold*

24

D. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dipaparkan sebelumnya, ditemukan bahwa penggunaan skala Likert 5 poin di Jr.MAI sudah berfungsi dengan baik. Reliabilitas Jr.MAI termasuk dalam kategori baik dan dapat mengklasifikasikan item dan person yang berasal lebih dari tiga kelompok. Sebanyak 18 item cocok dengan model dan bebas dari bias gender. Jadi Jr.MAI (*Junior Metacognitive Awareness Inventory*) versi B

memiliki sifat psikometrik yang baik untuk mengukur kesadaran metakognitif siswa SMA di Indonesia. Namun, peta person-item menunjukkan bahwa masih ada ruang untuk perbaikan untuk mencakup semua kemampuan siswa dengan membuat beberapa item yang sulit. Tidak ada bias substansial yang terdeteksi menurut probabilitas tanggapan dan ukuran DIF berdasarkan gender.

E. Potensi penelitian mendatang

Keterbatasan penelitian ini adalah belum dapat digunakan pada siswa yang berasal dari sekolah swasta karena responden yang terlibat berasal dari sekolah negeri. Atribut budaya siswa tidak dimasukkan untuk melihat apakah Jr.MAI bebas dari bias budaya. Namun, penelitian ini telah memberikan kontribusi yang signifikan dalam mengevaluasi sifat psikometrik Jr.MAI untuk digunakan di Indonesia dengan pendekatan item response theory. Temuan ini memiliki implikasi yang signifikan bagi guru, konselor, dan orang tua untuk membantu siswa mencapai kesuksesan akademik.

Saran untuk penelitian selanjutnya adalah fokus pada *Differential Item Functioning*. Oleh karena itu, penelitian selanjutnya perlu memperhatikan heterogenitas atribut responden. Misalnya, seorang peneliti dapat menilai bias Jr.MAI terhadap jenis sekolah (sekolah swasta dan negeri), lokasi sekolah (sekolah perkotaan dan pedesaan), atau berdasarkan bidang minat (bidang sains dan sosial).

F. Daftar Pustaka

- [1] R. Abdellah, "Metacognitive Awareness and its Relation to Academic Achievement and Teaching Performance of Pre-service Female Teachers in Ajman University in UAE," *Procedia - Soc. Behav. Sci.*, vol. 174, pp. 560–567, Feb. 2015, doi: 10.1016/j.sbspro.2015.01.707.
- [2] K. Craig, D. Hale, C. Grainger, and M. E. Stewart, "Evaluating Metacognitive Self-Reports: Systematic Reviews of the Value of Self-Report in Metacognitive Research," *Metacognition Learn.*, vol. 15, no. 2, pp. 155–213, Aug. 2020, doi: 10.1007/s11409-020-09222-y.

BAB 2

Validasi Instrumen Persepsi Belajar Online POSTOL

A. Perumusan Masalah

Pandemi Covid-19 telah mempercepat perubahan mendadak dan tidak terduga dalam cara belajar yang disukai siswa dan guru. Pendekatan sementara yang paling besar dalam menekan laju penularan Covid-19 di berbagai belahan dunia, termasuk Indonesia, adalah dengan menggunakan teknik pembelajaran daring/*online*. Mendukung metode pembelajaran *online* sangat penting untuk meminimalkan dampak pandemi terhadap pendidikan [1]–[3]. Media pembelajaran *online* telah meningkatkan kapasitas pengetahuan dan keterampilan siswa [4], [5]. Namun karena perubahan yang cepat dan tidak dapat diprediksi ini, siswa mungkin belum sepenuhnya siap untuk pembelajaran *online* [6], [7]. Menganalisis persepsi siswa terhadap pembelajaran *online* akan membantu guru dan pemangku kepentingan dalam mengembangkan kebijakan berikutnya.

Persepsi siswa terhadap pembelajaran *online* merupakan isu penting dalam pendidikan *online* [8]. Siswa memahami kelebihan dan kekurangan pembelajaran berbasis internet [9], [10]. Sikap yang baik terhadap pembelajaran *online* akan membantu integrasi dan keberhasilan proses

[11]. Di sisi lain, strategi pembelajaran *online* membebani siswa dan orang tua [12]–[14]. Dukungan sumber daya dan dorongan kebutuhan belajar mempengaruhi pembelajaran siswa selama pandemi [15]. Fasilitas tersebut antara lain ketersediaan perangkat keras dan perangkat lunak serta koneksi internet [16], [17], dan internet di Indonesia masih belum merata [8], [18].

Sejumlah penelitian telah meneliti persepsi siswa terhadap pembelajaran *online* selama dua tahun terakhir [19]–[23]. Salah satu skala yang dikembangkan untuk mengukur persepsi adalah Persepsi Siswa Terhadap Pembelajaran Online (*Perception of Students Towards Online Learning*, POSTOL). Skala ini dikembangkan oleh Bhagat et al. [24] dan telah diterapkan pada mahasiswa S1, S2, dan S3 di Taiwan. Salah satu skala yang dievaluasi untuk mengetahui perasaan siswa terhadap pembelajaran *online* adalah POSTOL. Melalui penggunaan teori tes klasik, kualitas POSTOL telah dievaluasi. POSTOL dibentuk oleh empat faktor: Kehadiran Sosial (*Social Presence*, SP), Karakteristik Instruktur (*Instructor Characteristics*, IC), Desain Instruksional (*Instructional Design*, ID), dan Kepercayaan (*Trust*, TR). Namun karena adanya disparitas tingkat pendidikan dan keadaan budaya, skala ini tidak bisa langsung diterapkan pada siswa di Indonesia. Oleh karena itu, diperlukan proses adaptasi untuk menilai atribut psikometrik POSTOL [25]–[27].

Evaluasi terhadap persepsi mahasiswa terhadap pembelajaran daring perlu segera dilakukan untuk melihat dukungan dan hambatan pelaksanaannya selama dua tahun terakhir. Hal ini bertujuan untuk meningkatkan efektivitas dan efisiensi pembelajaran daring. Sehingga, hal ini tidak menyebabkan hilangnya pembelajaran pada siswa. Hasil evaluasi akan memberikan informasi yang relevan dan akurat apabila menggunakan skala yang mempunyai sifat psikometrik yang baik.

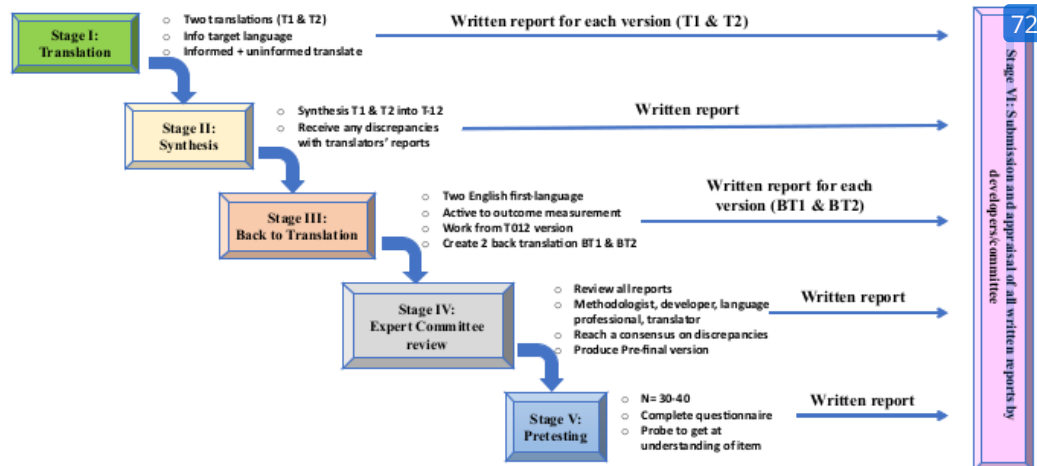
Hasil evaluasi dengan pendekatan teori tes klasik, EFA, dan CFA yang telah dilaporkan sebelumnya, belum memberikan informasi yang komprehensif mengenai sifat psikometri. Jadi, teori tes kontemporer harus digunakan untuk mendukungnya (model Rasch). Model Rasch memberikan rincian tambahan yang mendalam tentang sifat psikometrik. Contoh sifat psikometrik yang tidak dapat dijelaskan oleh teori tes klasik antara lain adalah berfungsinya skala penilaian Likert, unidimensi, bias skala terhadap demografi responden, dan kesesuaian item (tingkat kesulitan soal dan kemampuan responden) [28]. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi, dengan menggunakan model Rasch, sifat psikometri POSTOL pada siswa sekolah menengah dan lingkungan budaya Indonesia. Makalah ini menambah dan mendukung kemampuan psikometri POSTOL untuk beroperasi melintasi batas-batas budaya.

Teori: Proses adaptasi lintas budaya

Salah satu kesalahan umum dalam mengadaptasi alat ukur adalah hanya mengandalkan terjemahan dari bahasa asal ke bahasa tujuan. Lebih lanjut, proses adaptasi lebih dari sekedar menerjemahkan alat ukur. Namun, perlu adanya kontekstualisasi situasi sosio-kultural pengguna destinasi. Telah dipahami secara luas bahwa item-item tersebut harus diterjemahkan secara linguistik dan dikontekstualisasikan secara budaya jika alat pengukuran ingin digunakan lintas budaya. Item yang diterjemahkan secara linguistik dan budaya berupaya untuk menjunjung validitas konseptual isi instrumen dalam berbagai konteks budaya.

Skala laporan mandiri oleh Beaton *et al.* [29] diterjemahkan dan diadaptasi secara budaya mengikuti standar adaptasi lintas budaya. Standar yang diterima secara umum menyetujui versi final untuk pemilihan

alat ukur [26]. Proses adaptasi bertujuan untuk memastikan bahwa sumber dan target kuesioner setara secara semantik, idiomatis, berdasarkan pengalaman, dan secara konseptual. Prosedur yang disarankan untuk adaptasi lintas budaya ditunjukkan pada Gambar 6.



Gambar 6. Proses adaptasi lintas budaya

Teori: Skala POSTOL

POSTOL adalah salah satu skala Bhagat et al. [24] dikembangkan untuk mengevaluasi persepsi siswa terhadap pelaksanaan pembelajaran online. Akibat adanya pandemi Covid-19, hampir seluruh pembelajaran saat ini menggunakan modus daring, beralih dari modus tatap muka yang selama ini dilakukan. Salah satu strategi untuk menghentikan penyebaran virus Covid-19 adalah dengan menerapkan pendidikan online. Lebih dari dua tahun kegiatan pembelajaran berlangsung secara daring. Berbagai pengalaman dialami siswa selama pembelajaran daring. Hal ini menimbulkan persepsi yang berbeda-beda antar individu siswa.

Skala POSTOL yang dikembangkan oleh Bhagat et al. [24] terdiri dari empat dimensi: *Social Presence* (SP), *Instructor Characteristics* (IC), *Instruksional Design* (ID), dan *Trust* (TR). Keempat dimensi ini ditetapkan

melalui analisis faktor 2 tahap. Analisis faktor pertama dilakukan melalui *Exploratory Factor Analysis* (EFA). Semua item secara alami dikelompokkan berdasarkan data yang ada pada tahap ini. Tahap selanjutnya adalah struktur yang dibentuk pada tahap EFA, dikonfirmasi kembali melalui *Confirmatory Factor Analysis* (CFA), dan diperoleh 16 item yang memenuhi fit model.

Teori: Model Rasch

Model Rasch pertama kali dikembangkan oleh ahli matematika Denmark, Georg Rasch [30]. Pemodelan rasch merupakan bagian dari *Item Response Theory* (IRT) yang hanya berfokus pada satu parameter logistik yaitu tingkat kesulitan item yang dilihat dari dua sisi (tingkat kesulitan item dan kemampuan orang) [31]. Model Rasch dikembangkan untuk mengukur sifat-sifat laten manusia, seperti aspek kognitif dan non-kognitif (pendapat atau persepsi). Karena pengukuran merupakan variabel laten, maka model Rasch menempatkan posisinya sebagai model yang dapat mengubah instrumen menjadi skala ukur seperti alat ukur dalam fisika. Oleh karena itu, ide mendasar di balik model Rasch adalah membuat penggaris logit dengan skala interval yang sama, baik untuk tingkat kesulitan soal maupun kemampuan orang [32]. Model ini dapat membuat hierarki antara orang (peserta tes atau siswa) dan soal tes [33]. Model Rasch menggunakan model probabilistik. Siswa dapat memberikan respon yang akurat tergantung pada perbandingan kemampuan seseorang dan kesulitan soal. Skor mentah diproses menggunakan persamaan logaritmik untuk membandingkan kemampuan orang tersebut dan tingkat kesulitan item secara langsung.

Teori: Sifat Psikometri

Ketika memilih dan menggunakan instrumen untuk mengukur konstruksi yang tidak dapat diobservasi, penting untuk memeriksa sifat psikometriknya [34]. Validitas dan reliabilitas alat ukur disebut dengan sifat psikometrik [35]. Sebelum dapat dinyatakan bahwa kuesioner mempunyai ciri psikometrik yang baik, artinya dapat dipercaya dan valid, maka skala tersebut harus dianalisis secara menyeluruh [36]. Kegiatan utama dalam psikometri meliputi konstruksi atau penyusunan berbagai teori psikologi menjadi alat ukur psikologi/alat tes psikologi, serta pengembangan dan analisis data dari pengukuran tersebut [37].

Penyelidikan pengukuran kualitas seperti invariansi pengukuran, konsistensi internal, dan validitas struktural dalam pendidikan telah dilakukan secara luas menggunakan analisis Rasch sebagai pendekatan psikometri kontemporer [38]. Aspek yang diselidiki untuk mengevaluasi sifat psikometri meliputi (a) Reliabilitas orang dan item, indeks pemisahan orang dan item serta konsistensi internal, (b) Kesesuaian item dengan model dan tingkat kesulitannya, (c) analisis komponen utama (PCA) dari sisa untuk validitas struktural, dan (d) fungsi diferensial item (DIF) untuk mengukur invarian [28], [39]–[41].

B. Metodologi

Ukuran sampel harus ditentukan untuk menjamin stabilitas hasil estimasi. Diperlukan jumlah sampel minimal 50 orang untuk mencapai akurasi 1 logit dengan tingkat kepercayaan 99 persen [42]. Ling Lee et al. [43] menyarankan penggunaan antara 50 hingga 250 responden untuk mengevaluasi kebaikan model. Oleh karena itu, diyakini bahwa 176 responden memenuhi ukuran sampel minimum. Analisis tersebut tidak

memasukkan 3 dari 176 responden karena mereka berada dalam kondisi *outlier*. Tabel 4 mencantumkan informasi demografi responden.

Tabel 4. Statistik demografi

Demografi	Kategori	Jumlah	%
Gender	Laki-laki	95	54.0
	Perempuan	81	46.0
	Total	176	100.0
Tingkat sekolah	SMP	119	67.6
	SMA	57	32.4
	Total	176	100.0
Usia (tahun)	Rata-rata	13.5	
	SD	1.4	

POSTOL yang diterjemahkan ke dalam versi bahasa Indonesia berasal dari skala persepsi siswa terhadap pembelajaran *online* yang dikembangkan oleh Bhagat *et al.* [24]. POSTOL terdiri dari 4 faktor/dimensi yaitu: *Social Presence* (SP, 5 item), *Instructor Characteristics* (IC, 5 item), *Instructional Design* (ID, 3 item), dan *Trust* (TR, 3 item). Proses penerjemahan dilakukan oleh dosen bidang bahasa Inggris dengan menggunakan teknik *forward-backward translation* [44]. Survei *WhatsApp* dibuat dari terjemahan dan dikirim ke calon responden. Guru sekolah setempat berpartisipasi dalam proses pengumpulan data selama dua minggu. Peneliti menjamin kerahasiaan informasi yang diberikan oleh responden, dan partisipasi siswa bersifat opsional. Kami menekankan hal ini ketika memperkenalkan instrumen ini untuk memberikan fleksibilitas kepada responden dalam memberikan tanggapan.

Empat metrik utama—keandalan, kesesuaian model, penggunaan skala penilaian Likert 5 poin, dan unidimensi—digunakan untuk menilai sifat psikometrik POSTOL. Model Rasch digunakan untuk menguji sifat psikometrik instrumen. Untuk menganalisis data, digunakan *Winsteps* 4.6.1 dan *Ms.Excel*. Kami menggunakan nilai batas $\geq 0,70$ untuk

menunjukkan keandalan karena direkomendasikan [45]. Kesesuaian item terhadap model dievaluasi menggunakan kriteria *Infit MnSq* dan *Outfit MnSq* dengan rentang 0,5 – 1,5 [46]. Fungsionalitas skala penilaian Likert dievaluasi berdasarkan kriteria yang digunakan oleh Llamas-Ramos et al. [47]. Sedangkan unidimensionalitas dievaluasi berdasarkan *raw variance explained by measures* dan *unexplained variance in the 1st contrast*.

C. Hasil dan Pembahasan

Ringkasan Statistik POSTOL

Gambaran umum temuan statistik dari adaptasi instrumen POSTOL disajikan pada Tabel 5. Penelitian menunjukkan bahwa indeks pemisahan item dan orang adalah 8,00 dan 1,67. Reliabilitas orangnya adalah 0,74, dan reliabilitas itemnya adalah 0,98. Sedangkan skor reliabilitas tes yang ditunjukkan dengan nilai *Cronbach alpha* adalah 0,75.

Tabel 5. Ringkasan statistik POSTOL

	Separation	Reliability
Person	1.67	0.74
Item	8.00	0.98
Test	-	0.75

Karakteristik item

Tabel 6 merangkum indeks kecocokan dari 16 item di POSTOL berdasarkan entri. Berdasarkan Tabel 6, *Infit MnSq* berada pada rentang 0,74 hingga 1,45, sedangkan *Outfit MnSq* berada pada rentang 0,71 hingga 1,52. Analisis item menghasilkan tingkat kesulitan berkisar antara -1,65 hingga 1,19 logit, dan model *Standard Error* (S.E) berkisar antara 0,08-0,17 logit. ID3 merupakan item termudah dengan nilai Model S.E sebesar 0,17. Sebaliknya, soal tersulit dimiliki oleh TR2 dengan nilai Model

S.E sebesar 0,08. Nilai rata-rata butir soal adalah 0,00, dan standar deviasinya adalah 0,93.

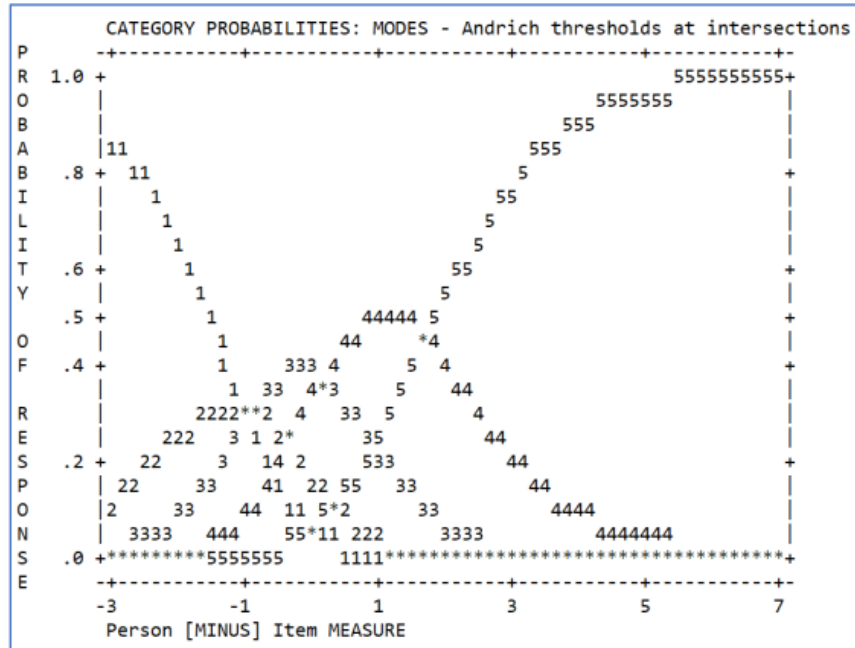
Tabel 6. Karakteristik item di POSTOL

Item	Measure	Model S.E	Infit MnSq	Outfit MnSq	Pt. Mea. Corr
IC1	-1.15	0.14	1.01	0.93	0.36
IC2	-0.71	0.12	1.00	0.94	0.38
IC3	-1.09	0.14	0.96	1.03	0.33
IC4	-0.76	0.13	0.86	0.92	0.43
IC5	-1.15	0.14	1.10	0.93	0.43
SP1	0.36	0.10	0.86	0.85	0.53
SP2	0.54	0.09	0.92	0.95	0.53
SP3	0.43	0.10	0.78	0.85	0.53
SP4	0.15	0.10	0.74	0.71	0.58
SP5	1.13	0.08	1.45	1.52	0.36
ID1	0.23	0.10	0.87	0.87	0.52
ID2	0.49	0.09	0.94	0.95	0.40
ID3	-1.65	0.17	1.04	0.93	0.28
TR1	0.45	0.09	1.20	1.16	0.55
TR2	1.19	0.08	1.24	1.25	0.44
TR3	1.54	0.08	1.00	1.01	0.52

Skala Peringkat Likert

Analisis skala penilaian dilakukan untuk membuktikan fungsionalitas skala peringkat Likert 5 poin yang digunakan di POSTOL. Tabel 7 menunjukkan sifat struktur skala penilaian Likert yang digunakan. Pada kolom kedua, kategori jawaban terbanyak berada pada kategori 5 (Sangat Setuju), 4 (Setuju), dan 3 (Ragu-ragu). Kolom ketiga menunjukkan rata-rata seluruh orang yang memilih setiap kategori. Rata-rata ini meningkat secara monoton. Pada kolom keempat dan kelima, Infit MnSq berada pada rentang 0,92 hingga 1,19, dan nilai Outfit MnSq sebesar 0,88 hingga 1,29 yang menunjukkan bahwa setiap kategori berada dalam batas yang dapat diterima. Kolom keenam menunjukkan perkiraan ambang batas POSTOL dalam urutan nol, -1,03, -0,86, 0,17, dan 1,72. Secara grafis respon masing-masing kategori direpresentasikan melalui kurva probabilitas pada

Gambar 7. Berdasarkan kurva probabilitas, skala kategori 2 tidak menunjukkan puncak tersendiri sehingga tidak mewakili satuan konstruksi yang diukur. Hal ini mengikuti nilai ambang batas pada Tabel 7.



Gambar 7. Kurva probabilitas skala peringkat Likert 5 poin POSTOL

Tabel 7. Statistik fungsionalitas skala peringkat Likert lima poin

Category	Rating scale	Count (%)	Observed Average	Infit MnSq	Outfit MnSq	Andrich Threshold
Sangat Tidak Setuju	1	60 (2)	-0.15	1.19	1.29	NONE
Tidak Setuju Ragu-	2	145 (5)	-0.01	0.92	0.88	-1.03
Ragu	3	470 (17)	0.59	1.01	1.01	-0.86
Setuju	4	1007 (36)	1.29	1.00	0.91	0.17
Sangat Setuju	5	1086 (39)	2.33	1.00	0.99	1.72

Unidimensi

Unidimensi skala POSTOL ditentukan melalui PCA residu. Secara empiris, ¹² *raw variance explained by measures* adalah 45,7%, *unexplained variance in the 1st contrast* adalah 10,0%, dan nilai Eigen (*Eigen Values*) adalah 2,78. Metrik ini diperlukan untuk mengetahui apakah POSTOL dapat mengukur persepsi siswa selama pembelajaran *online* secara akurat. Jika ³⁵ *raw variance explained by measures* lebih dari 40% dan *unexplained variance in the 1st contrast* kurang dari 15%, maka unidimensi skala tercapai [48], [49]. Hal ini menunjukkan bahwa POSTOL memiliki unidimensi yang baik.

Pembahasan

Penelitian ini bertujuan untuk mengevaluasi kualitas psikometri dari instrumen yang diadaptasi yang digunakan untuk mengukur persepsi siswa terhadap pembelajaran *online* selama Covid-19. Perangkat lunak Winsteps versi 4.6.1 digunakan untuk menganalisis data guna memverifikasi validitas konstruk POSTOL [50]. Hasil uji statistik awal menunjukkan bahwa orang tersebut dapat membedakan 16 item dalam 8 kelompok [51]. Linacre [52] menyatakan bahwa indeks pemisahan yang baik adalah $> 2,0$. Reliabilitas orang termasuk dalam kategori Baik, dan reliabilitas item termasuk dalam kategori sangat baik [45], [53]. Hal ini menunjukkan konsistensi jawaban responden, dan kualitas item di POSTOL tergolong istimewa. Sedangkan kualitas interaksi antara orang dengan item secara keseluruhan dilihat dari nilai *Cronbach alpha* [33]. Hasil analisis menunjukkan bahwa orang dan item mempunyai interaksi yang baik. Temuan ini mendukung hasil analisis reliabilitas instrumen POSTOL, seperti yang dievaluasi oleh Bhagat *et al.* [24].

Langkah selanjutnya adalah mempelajari secara cermat indeks kecocokan melalui *Infit MnSq* dan *Outfit MnSq*. Hasil analisis menunjukkan bahwa semua item cocok dengan model Rasch kecuali item SP5, “Membaca karya teman sekelas saya akan membantu meningkatkan kualitas pekerjaan saya.” Item SP5 memiliki nilai *Outfit MnSq* sebesar 1,52, di luar rentang 0,5-1,5. Namun, *Pt. Mea. Corr.* memiliki nilai 0,30-0,70 [54], [55]. *Pt. Mea. Corr.* tingkat yang tinggi menunjukkan bahwa suatu item dapat membedakan kemampuan responden [54]. Artinya, pola respon mempunyai orientasi searah dengan pola respon secara umum. Jadi item SP5 perlu dipertahankan. Ketidaksesuaian butir SP5 dapat berupa penggunaan kata-kata negatif atau memberikan kesan negatif [56]. POSTOL versi bahasa Indonesia dengan menggunakan model Rasch mendukung validitas POSTOL versi asli yang dianalisis menggunakan pendekatan analisis faktor [24].

Fungsionalitas skala Likert 5 poin dievaluasi berdasarkan ambang batas. Meskipun terjadi peningkatan nilai ambang batas dengan nilai kategori, namun peningkatan ambang batas tidak teratur. Hal ini menunjukkan bahwa kategori-kategori tersebut tidak didefinisikan secara jelas bagi responden. Responden tidak dapat membedakan dengan jelas 5 pilihan skala Likert yang disediakan, sehingga perlu menyederhanakan skala penilaian menjadi 4 skala penilaian Likert [57]. Gambar 7 memvisualisasikan penggabungan skala 2 dan 3 karena skala 2 tidak memiliki puncaknya sendiri. Sehingga penggunaan skala menjadi lebih efektif karena interval kategori menjadi lebih luas [56]. Temuan ini melengkapi sifat psikometri POSTOL yang belum pernah dilaporkan sebelumnya oleh Bhagat *et al.* [24].

Unidimensionalitas merupakan salah satu ukuran mendasar untuk menilai kemampuan ³³ suatu instrumen dalam mengukur apa yang akan

diukur [33], [58]. Berdasarkan nilai *raw variance explained by measures* dan *unexplained variance in the 1st contrast* menunjukkan bahwa 16 item pada instrumen POSTOL dapat diperlakukan sebagai ukuran unidimensi, dan tidak terdapat *noise* atau gangguan dalam pengukurannya. Secara lebih mendalam, ketidakhadiran item dari dimensi lain dieksplorasi melalui *Eigen value* (nilai Eigen) kurang dari 3. Dengan demikian, instrumen POSTOL yang diadaptasi memiliki unidimensi yang baik, dan tidak ditemukan indikasi gangguan (*noise*) dan item dari dimensi lain. Tidak terpenuhinya ukuran unidimensi dapat membahayakan reliabilitas dan validitas estimasi konstruk [58].

D. Kesimpulan

Sifat psikometrik instrumen POSTOL versi bahasa Indonesia dievaluasi berdasarkan model Rasch. Hasil analisis menunjukkan bahwa instrumen POSTOL mempunyai sifat psikometri yang baik untuk mengukur persepsi siswa terhadap pembelajaran *online* di Indonesia untuk SMP dan SMA. Secara statistic, instrumen POSTOL versi bahasa Indonesia memenuhi unsur validitas dan reliabilitas yang baik. Enam belas item yang diuji memenuhi unsur kesesuaian dengan model Rasch. Penggunaan skala penilaian Likert 4 poin lebih efektif untuk SMP dan SMA di Indonesia dibandingkan skala penilaian 5 poin pada versi aslinya. Selain itu, hasil uji unidimensi menunjukkan seluruh item pada instrumen POSTOL bahasa Indonesia memenuhi unsur unidimensi. Temuan ini merekomendasikan agar guru atau instruktur mengevaluasi persepsi siswa terhadap pembelajaran yang dilakukan selama pandemi Covid-19.

E. Potensi penelitian mendatang

Temuan ini harus dibatasi pada tingkat SMP dan SMA karena belum menjangkau berbagai demografi siswa. Penelitian di masa depan harus mengevaluasi sifat psikometrik instrumen dalam konteks yang lebih heterogen. Keberagaman jenis sekolah dan disiplin ilmu siswa (sosial, sains, kesehatan, atau kejuruan) perlu diperhatikan untuk memperoleh informasi pemanfaatannya dalam wilayah yang lebih luas. Kami merekomendasikan evaluasi sifat psikometrik POSTOL pada siswa tingkat sekolah dasar.

F. Daftar Pustaka

- [1] A. E. Clark, H. Nong, H. Zhu, and R. Zhu, "Compensating for Academic Loss: Online Learning and Student Performance During the COVID-19 Pandemic," *China Econ. Rev.*, vol. 68, no. May, p. 101629, 2021, doi: 10.1016/j.chieco.2021.101629.
- [2] M. B. Ulla and W. F. Perales, "Facebook as an Integrated Online Learning Support Application During the COVID19 Pandemic: Thai university students' Experiences and Perspectives," *Heliyon*, vol. 7, no. 11, p. e08317, Nov. 2021, doi: 10.1016/j.heliyon.2021.e08317.
- [3] N. Zafar and J. Ahamed, "Emerging Technologies for the Management of COVID19: A Review," *Sustain. Oper. Comput.*, vol. 3, no. May, pp. 249–257, 2022, doi: 10.1016/j.susoc.2022.05.002.
- [4] R. Chaker, F. Bouchet, and R. Bachelet, "How Do Online Learning Intentions Lead to Learning Outcomes? The Mediating Effect of the Autotelic Dimension of Flow in a MOOC," *Comput. Human Behav.*, vol. 134, p. 107306, Sep. 2022, doi: 10.1016/j.chb.2022.107306.
- [5] A. Hurajova, D. Kollarova, and L. Huraj, "Trends in Education During the Pandemic: Modern Online Technologies as a Tool for the Sustainability of University Education in the Field of Media and Communication Studies," *Heliyon*, vol. 8, no. 5, p. e09367, May 2022, doi: 10.1016/j.heliyon.2022.e09367.
- [6] M. Maqableh and M. Alia, "Evaluation Online Learning of Undergraduate Students Under Lockdown Amidst COVID-19 Pandemic: The Online Learning Experience and Students' Satisfaction," *Child. Youth Serv. Rev.*, vol. 128, no. July, p. 106160, 2021, doi: 10.1016/j.childyouth.2021.106160.
- [7] Y. M. Tang et al., "Comparative Analysis of Student's Live Online Learning

BAB 3

Validasi Instrumen Model Penilaian E-Learning HELAM

A. Perumusan Masalah

Sekolah di berbagai jenjang pendidikan telah mengubah cara pembelajaran di masa pandemi Covid-19, termasuk sekolah-sekolah di Indonesia. Sebagian besar sekolah telah mengalihkan kegiatan pembelajarannya dari tatap muka menjadi daring/*online* seluruhnya atau sebagian [1], [2]. Hal ini diterapkan untuk mengantisipasi *learning loss* [3]. *Learning Management System* (LMS) menjadi salah satu *platform* pembelajaran yang dipilih karena pengelolaannya yang mudah. LMS memiliki pengelolaan materi yang terintegrasi, pemantauan peserta dan guru, serta penyesuaian proses belajar mengajar [4]. Pemanfaatan LMS dalam pembelajaran juga menjadi salah satu solusi untuk meminimalisir penyebaran virus di Indonesia dan dunia. Pandemi Covid-19 mempercepat penggunaan LMS yang merupakan salah satu *platform* pembelajaran ideal bagi guru di berbagai negara, termasuk di Indonesia.

Seluruh sekolah pada berbagai jenjang, di Indonesia dan di dunia, telah dipaksa untuk mengubah mode pembelajaran saat pandemic covid-19. Banyak sekolah yang menggeser kegiatan pembelajaran dari *face to*

face menjadi sepenuhnya atau sebagian *online* [5]–[8]. Hal ini dilakukan untuk mengantisipasi terjadinya *loss learning* [9]–[12].

LMS menjadi salah satu mode yang dipilih karena pengelolaannya yang mudah [13], [14]. LMS telah mengintegrasikan pengelolaan materi, pemantauan peserta dan guru, serta menyesuaikan proses belajar dan mengajar [15]–[17]. Penggunaan LMS dalam pembelajaran juga telah menjadi salah satu solusi untuk memutus mata rantai penyebaran virus di Indonesia dan di dunia [18]–[20]. Sehingga, secara sadar maupun tidak, pandemik Covid-19 telah mengakselerasi penggunaan LMS di bidang Pendidikan pada berbagai negara, termasuk Indonesia.

Siswa membutuhkan waktu adaptasi yang cukup agar merasa nyaman menggunakan LMS sebagai salah satu mode belajar saat ini. LMS telah berkontribusi secara signifikan bagi Pendidikan [21]–[23]. Namun, perubahan mendadak yang terjadi sejak akhir 2019 tidak sepenuhnya mendapat respon positif dari siswa [24]. Oleh karena itu, evaluasi berbagai kendala maupun pendukung pelaksanaan LMS selama 3 tahun terakhir di Indonesia perlu dilakukan. Bahkan menjadi persyaratan penting sebagai umpan balik dalam rangka perbaikan secara berkelanjutan [25].

Salah satu instrumen yang dapat digunakan dalam proses evaluasi *blended learning* adalah instrumen *Hexagonal E-Learning Assessment Model* (HELAM) yang dikembangkan oleh Ozkan dan Koseler [26]. Skala HELAM dapat menjadi instrumen alternatif dalam evaluasi pembelajaran di LMS. Persepsi pengguna terhadap LMS selama pandemi COVID-19 di Indonesia perlu diteliti. Sementara pengembangan instrumen baru membutuhkan proses yang panjang. Oleh karena itu, diperlukan proses adaptasi. HELAM merupakan skala evaluasi penerapan LMS di Indonesia yang sistematis dan komprehensif karena dimensi yang digunakan dalam HELAM memadukan aspek sosial dan teknis. Mengingat skala HELAM

dikembangkan dalam konteks budaya yang berbeda, maka diperlukan proses adaptasi agar dapat digunakan dalam konteks Indonesia [27].

Adaptasi HELAM pada konteks budaya yang berbeda, termasuk di Indonesia, kunci untuk menjamin validitas, keberlakuan, dan relevansinya, serta menghindari bias budaya dalam hasil penelitian dan validitas data. Proses adaptasi juga memastikan pemahaman dan respon yang konsisten dari responden dengan latar belakang budaya yang beragam, menjaga keterbukaan terhadap keragaman budaya, dan memungkinkan hasil penelitian lebih berlaku dan relevan dalam berbagai konteks budaya. Oleh karena itu, adaptasi HELAM menjadi tahap penting dalam menghasilkan penelitian yang berkualitas dan relevan di berbagai konteks budaya. Skala HELAM telah digunakan di Indonesia sebelumnya [28]. Namun, ada keterbatasan penelitian yang melaporkan mengenai kualitas skala HELAM pada konteks budaya Indonesia.

Kajian kualitas skala HELAM dengan menggunakan analisis faktor pada mahasiswa sarjana dan pasca sarjana di Brunel University, Inggris menunjukkan bahwa 65 item yang tersebar di enam dimensi mempunyai kualitas yang memuaskan yaitu *quality of system, service, content, learner perspective, instructor attitudes*, dan *supportive issues* [29]. Analisis skala HELAM dengan pendekatan analisis faktor mempunyai beberapa kelemahan [29]. Analisis faktor berdasarkan *Classical Test Theory* (CTT) hanya menggambarkan kualitas skala pada level instrumen, bukan pada level item individual. Keandalan seringkali terbatas pada *Cronbach's alpha*. Teori tes klasik memiliki sifat ketergantungan item dan ketergantungan orang. Masalah lain yang tidak dapat dipecahkan oleh CTT adalah dengan melihat fungsionalitas skala penilaian Likert yang digunakan di HELAM. Oleh karena itu, diperlukan analisis alternatif untuk

mengevaluasi skala HELAM pada studi kasus tertentu, dalam hal ini konteks budaya di suatu daerah.

Model Rasch merupakan analisis alternatif yang dapat mengatasi keterbatasan CTT. Dalam model Rasch, data mentah yang diperoleh diubah menjadi data interval [30]. Selain itu, analisis Rasch menghasilkan kesalahan standar yang lebih kecil dan memungkinkan estimasi reliabilitas dan validitas yang tidak bergantung pada properti *item-dependent* dan *person-dependent* [31], [32]. Hasil analisis model Rasch mampu menjelaskan kualitas skala hingga level individual item [33], [34]. Rasch dapat digunakan untuk melihat konsistensi pola respon yang ada. Analisis Rasch memberikan alternatif yang sangat efektif untuk mengeksplorasi sifat psikometrik dan memperhitungkan bias respon [35]. Lebih lanjut, model Rasch mampu mengevaluasi fungsionalitas skala penilaian Likert yang digunakan dalam skala tersebut [36]. Lebih penting lagi, model Rasch adalah teknik psikometri yang dapat meningkatkan akurasi pembuatan instrumen, memadukan kualitas instrumen, dan menilai kinerja responden [37]. Oleh karena itu, penelitian ini bertujuan untuk memancarkan sifat psikometri skala HELAM secara komprehensif pada konteks budaya Indonesia menggunakan pemodelan Rasch. Penelitian ini telah mengisi adanya evaluasi sifat psikometri skala HELAM untuk program sertifikasi guru di Indonesia.

B. Metodologi

Implementasi skala HELAM versi adaptasi telah dilakukan untuk mengevaluasi sifat psikometrik dalam konteks Indonesia. Responden yang terlibat adalah 326 orang guru yang mengikuti program TPE program jabatan di Universitas Ahmad Dahlan, Indonesia. Responden berasal dari 14 provinsi di Indonesia. Lokasi dipilih berdasarkan jumlah data yang

dibutuhkan untuk memenuhi ukuran sampel. Ukuran sampel minimal 50 responden dapat digunakan untuk memenuhi kecukupan stabilitas data dengan akurasi 1 Logit dengan tingkat kepercayaan 99% [38]. Sedangkan penggunaan ukuran sampel pada kisaran 250 dapat meningkatkan akurasi pengukuran hingga 0,5 logit pada tingkat kepercayaan 99% [39]. Oleh karena itu, ukuran sampel yang digunakan diharapkan berdampak pada stabilitas data dan model yang baik dalam analisis Rasch [40]. Karakteristik responden dirangkum pada Tabel 8.

Tabel 8. Karakteristik responden

	Frekuensi	Persentase (%)
Gender		
Perempuan	258	79.1%
Laki-laki	68	20.9%
Usia		
≤ 30	55	16.9%
31-35	106	32.5%
36-40	104	31.9%
≥ 41	61	18.7%
Mean		36.2
SD		6.0

Skala HELAM versi adaptasi telah dievaluasi kesesuaiannya dengan konteks Indonesia oleh 3 ahli bahasa melalui prosedur adaptasi [41], [42]. Skala HELAM terdiri dari 73 item yang dibagi menjadi 2 bagian, bagian pertama berupa data demografi untuk menjangkau data responden secara umum dan bagian kedua berupa pengalaman LMS Pembelajaran. Skala HELAM terdiri dari 6 dimensi: Perspektif Pembelajar (*Learner's Perspective, LP*), Sikap Instruktur (*Instructor Attitudes, IA*), Kualitas Sistem (*System Quality, SyQ*), Kualitas Konten Informasi (*Information Content Quality, ICQ*), Kualitas Layanan (*Service Quality, SeQ*), dan Masalah Pendukung (*Supportive Issues, SI*) [29]. Bagian kedua menggunakan skala Likert 4 poin, dari 1 (Sangat Tidak Setuju) hingga 4 (Sangat Setuju).

Kerahasiaan tanggapan peserta dijaga dan peserta diberi hak untuk menarik tanggapannya jika merasa tidak nyaman. Semua data dikumpulkan secara ⁷⁴ online dengan menggunakan Google Forms. ⁶⁹ Pengumpulan data dilakukan selama kurang lebih empat bulan dengan menggunakan teknik *convenience sampling*.

Model Rasch digunakan untuk mengevaluasi sifat psikometrik skala HELAM. Model ini dianggap sebagai perpanjangan atau alternatif dari CTT [43]. Model Rasch dipilih karena memungkinkan proses transformasi data dari ordinal ke interval untuk memenuhi asumsi HELAM ³⁴ sebagai alat ukur. Alat ukur yang baik harus mempunyai skala dengan jarak interval yang sama dengan alat ukur fisik pada umumnya. Untuk analisis data, kami menggunakan perangkat lunak Winsteps 6.4.1 [44].

Sebelum mengevaluasi sifat psikometrik item-item di HELAM, dilakukan proses *screening* awal terhadap orang dan skala penilaian yang digunakan. Data orang ekstrim dan orang yang memiliki *Pt. Mea. Corr.* negatif (ketidaksesuaian) dihilangkan dari analisis. *Pt. Mea. Corr.* negatif atau nol menunjukkan bahwa data mempunyai orientasi yang berlawanan dengan prediksi model. Setelah menyaring orangnya, dilanjutkan dengan penyaringan skala penilaian yang digunakan. Fungsionalitas skala penilaian disaring secara visual berdasarkan kurva probabilitas skala penilaian. Evaluasi sifat psikometrik skala HELAM diawali dengan analisis ¹ *Unidimensionality, Reliability, Item fit*, dan kemudian *Differential Item Functioning* (DIF). Reliabilitas ditinjau berdasarkan indeks pemisahan (*separation*) dan reliabilitas pada item, orang, dan *Cronbach's alpha*. ¹ Peta Wright mendukung validitas konstruk. *Item fit* menggunakan kriteria *Outfit MNSQ, Outfit ZSTD* dan *Pt. Mea. Corr.* Terakhir, *Differential Item Functioning* (DIF) digunakan untuk melihat bias item di HELAM terhadap gender.

C. Hasil dan Pembahasan

Penyaringan Awal

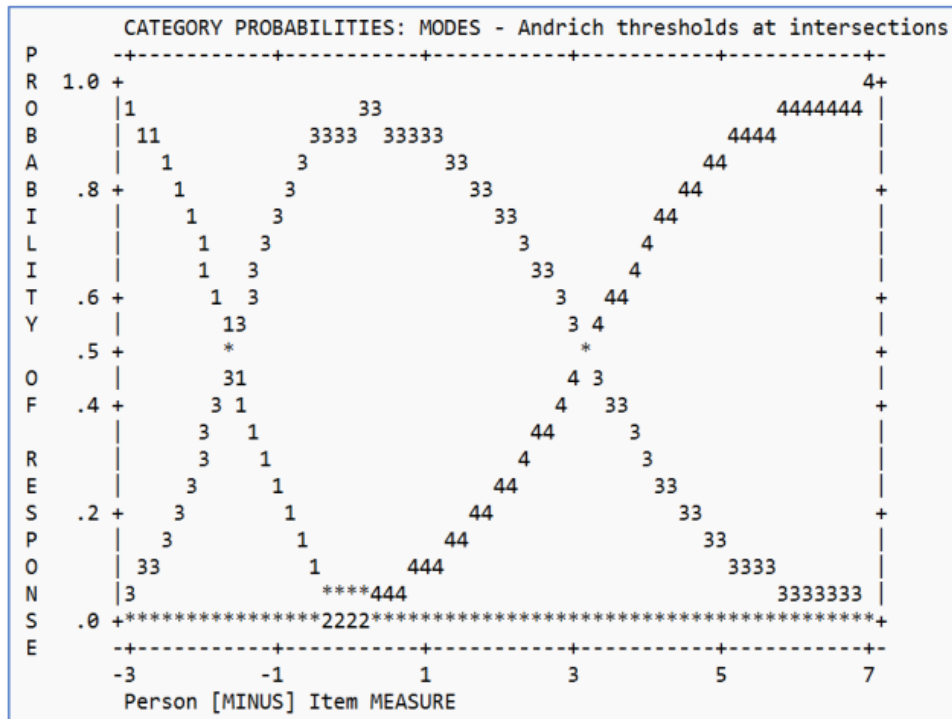
Penyaringan awal dilakukan untuk mendapatkan kualitas data yang baik. Penyaringan awal dilakukan terhadap orang dan skala penilaian yang digunakan. Dalam penyaringan, 82 (25%) orang dikeluarkan dari analisis karena *outlier* dan ketidakcocokan terhadap model. Hasil penyaringan skala penilaian disajikan pada Tabel 9.

Tabel 9. Fungsi Skala Rating pada HELAM

Category Label	Count	Observed Average	Infit MNSQ	Outfit MNSQ	Andrich Threshold
1 (SD)	665	-0.56	1.13	1.30	None
2 (D)	0	0	0.00	0.00	Null
3 (A)	12187	0.77	0.98	0.92	-3.18
4 (SA)	3008	3.06	0.94	0.86	3.18

SD = Sangat Tidak Setuju; D = Tidak Setuju; A = Setuju; SA = Sangat setuju

Tabel 9 menunjukkan bahwa nilai rata-rata teramati meningkat, dari -0,56 menjadi 3,06. Demikian pula, *Andrich Threshold* telah meningkat dari *None* menjadi 3,18. Hitungan yang diamati menunjukkan adanya peningkatan. Penggunaan skala penilaian juga menunjukkan fungsionalitas yang baik dari skala likert poin 1 (Sangat Tidak Setuju), 3 (Setuju), dan 4 (Sangat Setuju). Namun setiap indeks pada skala likert dengan rating 2 (Tidak Setuju) menunjukkan fungsionalitas yang belum optimal. Fungsionalitas skala Likert rating 2 (Tidak Setuju) divisualisasikan pada Gambar 8.



Gambar 8. Kurva Probabilitas skala Likert HELAM

Gambar 8 menunjukkan bahwa peringkat skala Likert 2 (Tidak Setuju) tidak memiliki puncak tersendiri sedangkan peringkat skala Likert lainnya menunjukkan fungsionalitas yang baik. Hal ini terlihat dari adanya masing-masing puncak pada skala Likert dengan rating 1 (Sangat Tidak Setuju), 3 (Setuju), dan 4 (Sangat Setuju).

Unidimensi

Unidimensi HELAM ditentukan dengan menggunakan Analisis Komponen Utama Residual atau *Principal Component Analysis Residual* (PCAR). Hasil evaluasi unidimensi HELAM ditinjau berdasarkan Varians mentah yang dijelaskan oleh ukuran dan Varians yang tidak dapat dijelaskan pada kontras pertama. Skor varians mentah yang dijelaskan oleh ukuran adalah 31,6% (nilai yang diharapkan adalah 34,0%) dan varians yang tidak dapat dijelaskan dalam skor kontras pertama adalah

4,0%. Sedangkan nilai Eigen pada Unexplained variance pada kontras pertama adalah 3,78.

Ringkasan Statistik

Setelah melakukan screening awal, dilakukan evaluasi reliabilitas skala HELAM. Keandalan ditinjau berdasarkan perspektif item dan orang. Hasil analisis reliabilitas berdasarkan parameter Rasch disajikan pada Tabel 10.

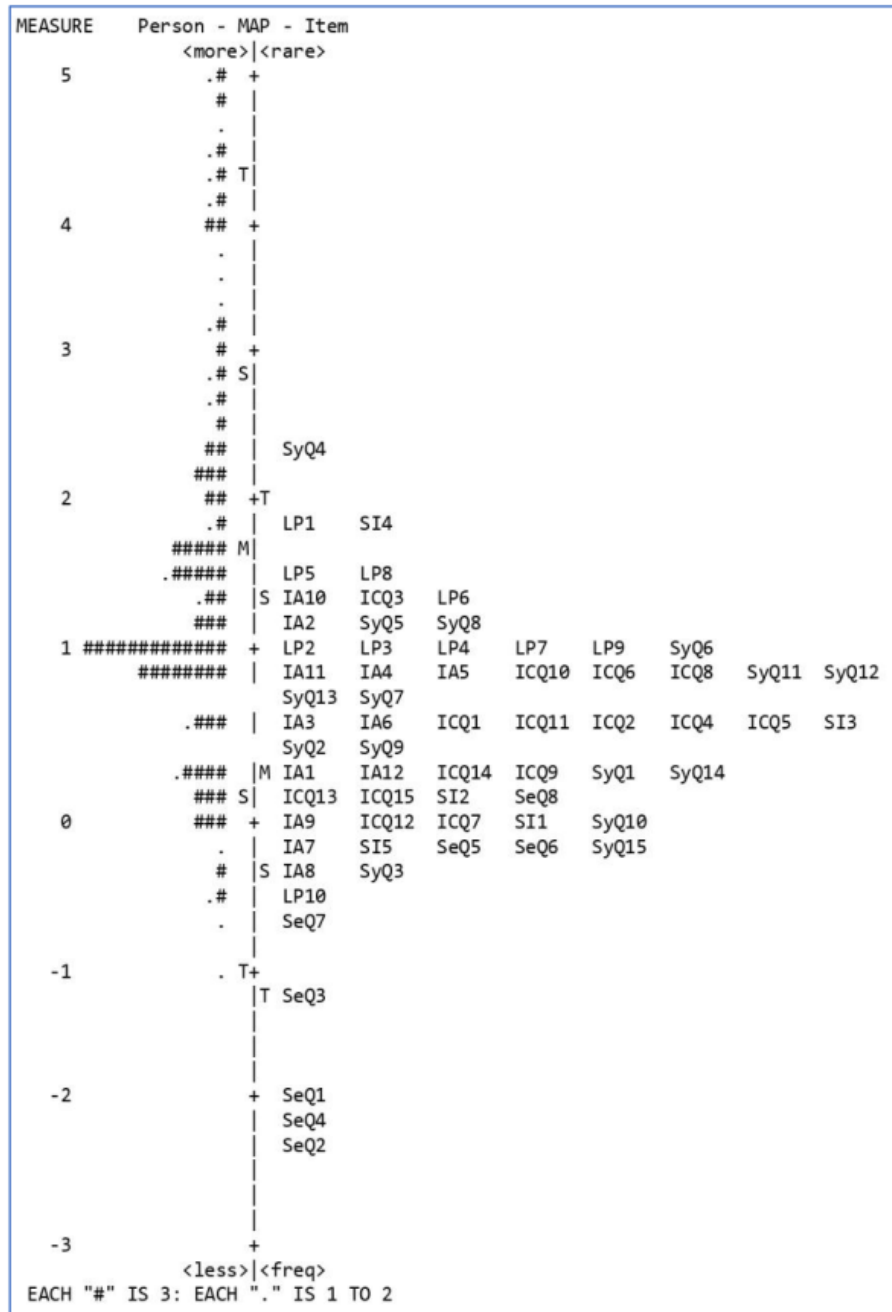
Tabel 10. Ringkasan Statistik HELAM dalam konteks Indonesia

	N	Mean	SD	Separation	Reliability	Cronbach's Alpha
Item	65	0.00	0.82	5.37	0.97	0.95
Person	244	1.15	1.34	4.12	0.94	

Tabel 10 menunjukkan bahwa skala HELAM dalam konteks Indonesia mempunyai reliabilitas yang baik. Reliabilitas item dan orang masing-masing sebesar 0,97 dan 0,94. Sedangkan reliabilitas instrumen ditunjukkan dengan indeks *Cronbach's Alpha* yaitu sebesar 0,95. Indeks pemisahan pada item dan orang menunjukkan reliabilitas yang baik, dengan skor 5,37 dan 4,12. Secara statistik rata-rata dan simpangan baku item adalah 0,00 logit dan 0,82 logit, sedangkan rata-rata dan simpangan baku orang adalah 1,15 logit dan 1,34 logit.

Wright map (Peta Wright)

Distribusi tingkat kesulitan soal HELAM dievaluasi secara visual menggunakan peta Wright. Gambar 9 memvisualisasikan distribusi tingkat kesulitan item dalam ukuran logit.



Gambar 9. Wright Map dari HELAM

Berdasarkan Gambar 9, logit bar membentang dari -3 logit hingga 5 logit. Rata-rata logit orang lebih tinggi dari rata-rata logit item. Hal ini menunjukkan bahwa rata-rata tingkat penerimaan LMS di kalangan guru lebih tinggi dibandingkan rata-rata tingkat kesukaran soal. Item logit berkisar dari -2,66 logit hingga 1,85 logit, dan logit orang berkisar dari -1,47 logit hingga 4,63 logit. Item yang paling mudah disetujui responden adalah item SeQ2 dan item yang paling sulit disetujui adalah item SyQ4.

Kesesuaian item (*Item fit*)

Kualitas item-item di HELAM ditinjau berdasarkan kesesuaiannya dengan nilai *infit MNSQ* dan *outfit MNSQ*. Kecocokan item dirangkum dalam Tabel 11. Tabel 11 menunjukkan bahwa terdapat 19 item pada skala HELAM yang perlu direvisi atau dihilangkan dari skala HELAM. Lima item (SyQ4, SeQ1, SeQ2, SeQ3, SeQ4) yang tersebar pada dimensi Kualitas Sistem dan Kualitas Layanan memiliki lokasi outlier. Lima belas item diidentifikasi sebagai ketidaksesuaian. Tiga item (LP1, LP5, LP10) pada dimensi Perspektif Peserta didik, tiga item pada dimensi Kualitas Sistem (SyQ1, SyQ4, SyQ12), lima item (ICQ1, ICQ2, ICQ3, ICQ4, ICQ9) pada dimensi Kualitas Konten Informasi, dan masing-masing 2 item pada dimensi Kualitas Pelayanan (SeQ5, SeQ6). Dimensi Masalah Pendukung (SI4, SI5) memiliki indeks *outfit MNSQ* di luar rentang penerimaan (0,50 – 1,50). Empat item (LP1, SyQ4, ICQ3, SI4) dari lima belas item *misfit* memiliki nilai *Pt. Mea. Corr.* di luar rentang penerimaan (0,40 – 0,85). Kecocokan item kemudian diperkuat dan didukung secara visual melalui Gambar 10.

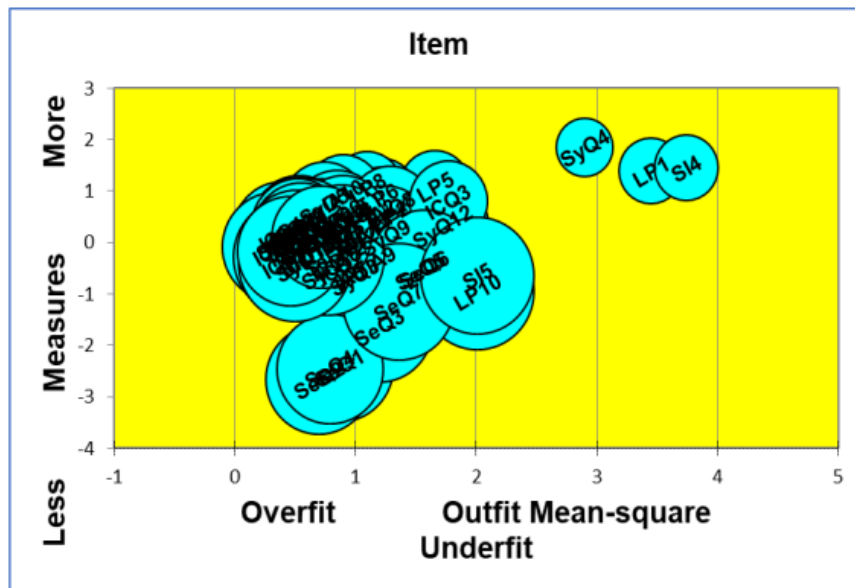
Tabel 11.1. Item fit HELAM dalam konteks Indonesia

Item	Mea. (logit)	S.E.	Diff. level	Item Fit Index		DIF Index		
				Outfit MNSQ	Outfit ZSTD	Cont	t	Prob.
LP1 ^b	1.39	0.09	very difficult	3.45	9.90	-0.48	-2.28	0.06
LP2	0.53	0.12	difficult	1.29	1.40	-0.02	-0.07	0.81
LP3	0.42	0.13	difficult	0.95	-0.19	-0.60	-2.14	0.19
LP4	0.47	0.13	difficult	0.93	-0.29	-0.20	-0.67	0.88
LP5 ^b	1.09	0.10	very difficult	1.66	3.38	-0.10	-0.43	0.69
LP6	0.87	0.11	very difficult	1.20	1.11	0.06	0.21	0.90
LP7	0.42	0.13	difficult	1.02	0.16	-0.17	-0.57	0.01
LP8	1.07	0.10	very difficult	1.10	0.63	-0.29	-1.27	0.36
LP9	0.42	0.13	difficult	0.83	-0.80	-0.36	-1.21	0.41
LP10 ^{bc}	-0.95	0.16	very easy	2.01	5.09	1.25	3.37	0.03
IA1	-0.05	0.15	easy	0.61	-2.10	0.26	0.70	0.05
IA2	0.61	0.12	difficult	1.21	1.08	0.56	1.70	0.07
IA3	0.10	0.14	difficult	0.86	-0.64	0.31	0.87	0.66
IA4	0.30	0.14	difficult	0.88	-0.54	-0.03	-0.08	0.28
IA5	0.34	0.13	difficult	0.81	-0.93	-0.29	-0.94	0.98
IA6	0.08	0.15	difficult	0.97	-0.08	0.69	1.87	0.81
IA7	-0.59	0.16	easy	0.98	-0.07	-0.15	-0.40	0.88
IA8	-0.66	0.16	easy	0.95	-0.24	-0.25	-0.66	0.68
IA9	-0.31	0.16	easy	1.21	1.04	0.05	0.14	0.87
IA10	0.93	0.11	very difficult	0.90	-0.50	-0.13	-0.51	1.00
IA11	0.24	0.14	difficult	1.13	0.65	-0.60	-2.00	0.58
IA12	0.02	0.15	difficult	1.02	0.19	-0.29	-0.84	0.81
SyQ1 ^b	0.02	0.15	difficult	0.38	-3.86	-0.17	-0.49	0.44
SyQ2	0.12	0.14	difficult	0.58	-2.30	-0.15	-0.44	0.17
SyQ3	-0.66	0.16	easy	0.99	0.00	0.03	0.09	0.71

Item	Mea. (logit)	S.E.	Diff. level	Item Fit Index			DIF Index		
				Outfit MNSQ	Outfit ZSTD	Pt. Mea. Corr.	Cont	t	Prob.
SyQ4 ^{ab}	1.85	0.08	very difficult	2.90	9.74	0.20	0.33	1.59	0.30
SyQ5	0.71	0.12	difficult	0.74	-1.46	0.50	-0.05	-0.17	0.79
SyQ6	0.49	0.13	difficult	0.85	-0.72	0.53	0.00	0.00	0.28
SyQ7	0.24	0.14	difficult	0.69	-1.57	0.53	-0.11	-0.32	0.45
SyQ8	0.64	0.12	difficult	1.29	1.44	0.40	0.10	0.34	0.46
SyQ9	0.10	0.14	difficult	1.25	1.18	0.60	-0.04	-0.13	0.61
SyQ10	-0.33	0.16	easy	0.58	-2.43	0.66	-0.11	-0.28	0.31
SyQ11	0.30	0.14	difficult	0.53	-2.66	0.54	0.31	0.89	0.69
SyQ12 ^b	0.32	0.13	difficult	1.72	2.93	0.43	0.00	0.00	0.23
SyQ13	0.39	0.13	difficult	0.90	-0.44	0.40	0.78	2.19	0.27
SyQ14	-0.03	0.15	easy	0.62	-2.02	0.53	0.29	0.78	0.56
SyQ15	-0.48	0.16	easy	0.80	-1.07	0.56	0.71	1.87	0.56
ICQ1 ^b	0.12	0.14	difficult	0.44	-3.32	0.60	0.47	1.29	0.07
ICQ2 ^b	0.10	0.14	difficult	0.42	-3.49	0.50	0.19	0.52	0.01
ICQ3 ^b	0.81	0.11	very difficult	1.77	3.47	0.19	0.40	1.34	0.06
ICQ4 ^b	0.18	0.14	difficult	0.39	-3.76	0.55	0.04	0.11	0.25
ICQ5	0.12	0.14	difficult	0.89	-0.47	0.54	0.07	0.20	0.63
ICQ6	0.28	0.14	difficult	0.53	-2.69	0.56	-0.16	-0.51	0.53
ICQ7	-0.46	0.16	easy	0.88	-0.57	0.56	-0.13	-0.34	0.38
ICQ8	0.24	0.14	difficult	0.59	-2.22	0.58	-0.60	-2.00	0.63
ICQ9 ^b	-0.09	0.15	easy	0.34	-4.29	0.68	0.20	0.55	0.16
ICQ10	0.32	0.13	difficult	0.52	-2.78	0.53	-0.59	-1.99	0.48
ICQ11	0.16	0.14	difficult	0.55	-2.51	0.60	-0.32	-0.97	0.97
ICQ12	-0.33	0.16	easy	0.86	-0.65	0.61	0.18	0.47	0.82
ICQ13	-0.26	0.16	easy	0.61	-2.16	0.62	0.00	0.00	0.80
ICQ14	-0.05	0.15	easy	0.51	-2.77	0.62	0.40	1.07	0.61

Item	Mea. (logit)	S.E.	Diff. level	Item Fit Index		DIF Index			
				Outfit MNSQ	Outfit ZSTD	Pt. Mea. Corr.	Cont	t	Prob.
ICQ15	-0.28	0.16	easy	0.50	-3.36	0.69	0.53	1.39	0.14
SeQ1 ^a	-2.42	0.15	very easy	0.87	-1.06	0.55	-0.28	-0.78	0.29
SeQ2 ^a	-2.66	0.15	very easy	0.70	-2.33	0.56	-0.19	-0.54	0.80
SeQ3 ^a	-1.64	0.15	very easy	1.20	1.64	0.54	-0.07	-0.20	0.82
SeQ4 ^a	-2.45	0.15	very easy	0.79	-1.75	0.54	-0.56	-1.57	0.03
SeQ5 ^b	-0.54	0.16	easy	1.54	2.57	0.53	0.05	0.13	0.31
SeQ6 ^b	-0.51	0.16	easy	1.57	2.64	0.51	-0.04	-0.10	0.92
SeQ7	-1.15	0.16	very easy	1.36	2.33	0.58	0.29	0.76	0.51
SeQ8	-0.28	0.16	easy	0.77	-1.16	0.61	0.24	0.64	0.84
SI1	-0.41	0.16	easy	0.50	-3.13	0.68	0.06	0.17	0.40
SI2	-0.16	0.15	easy	0.50	-3.25	0.63	0.11	0.31	0.78
SI3	0.10	0.14	difficult	0.72	-1.39	0.56	0.45	1.22	0.30
SI4 ^b	1.45	0.09	very difficult	3.74	9.90	0.18	-0.04	-0.17	0.85
SI5 ^b	-0.64	0.16	easy	2.01	4.44	0.50	0.06	0.16	0.51

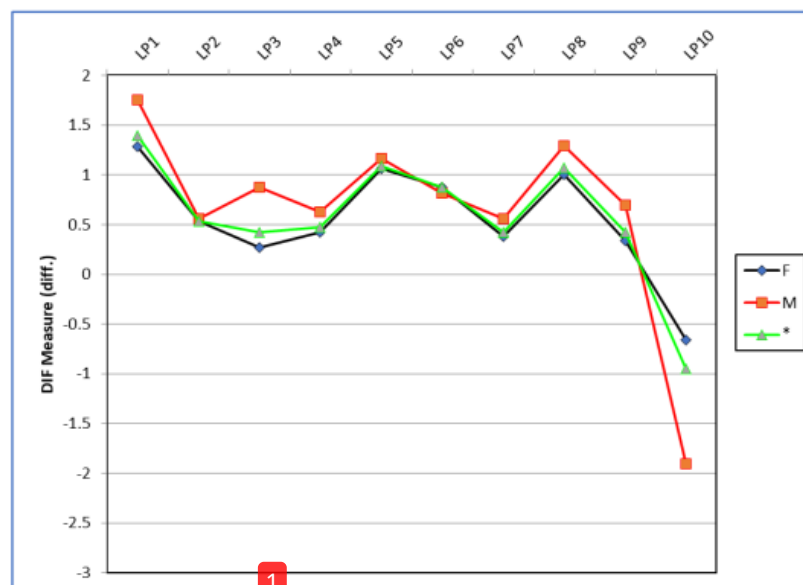
Remarks: ^a = item outlier; ^b = item misfit; c = item DIF; Mea. = Measure; S.E = Standard Error; Diff. level = Difficulty level; Cont. = Contrast; Prob. = Probability



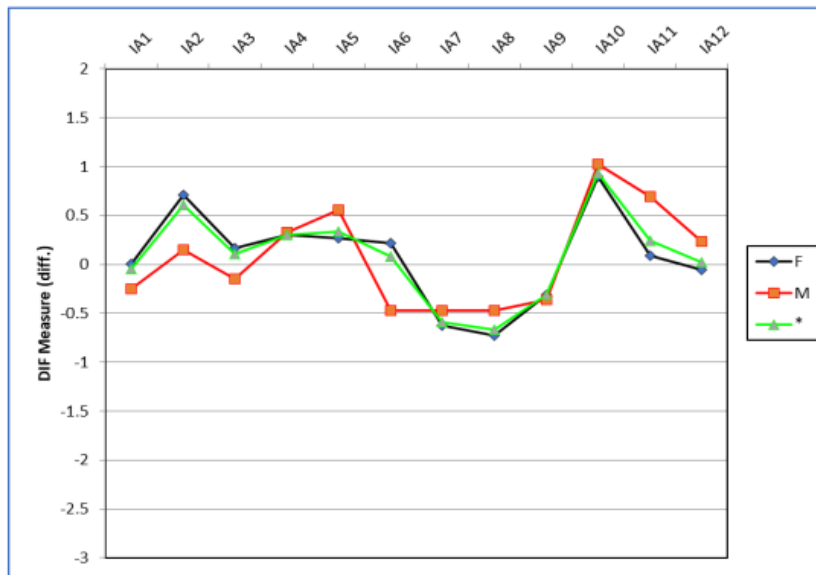
Gambar 10. Bagan Buble-fit item berdasarkan statistik Fit

1 Differential Item Functioning (DIF)

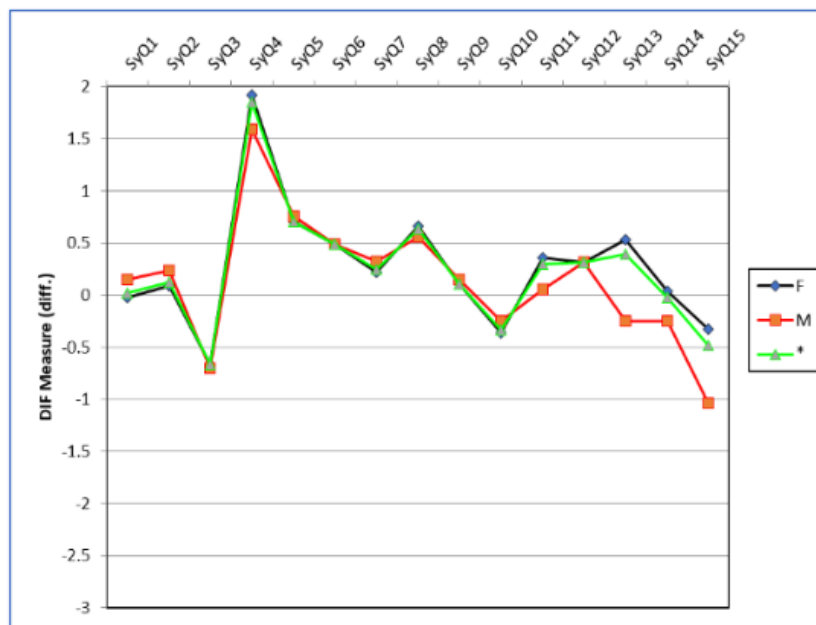
Bias item di HELAM dievaluasi melalui Differential Item Functioning (DIF). Indeks Kontras DIF, Rasch-Welch t, dan probabilitas Mantel Haenszel. Hasil analisis DIF sebanyak 65 item disajikan pada Tabel 11 dan divisualisasikan pada Gambar 1.



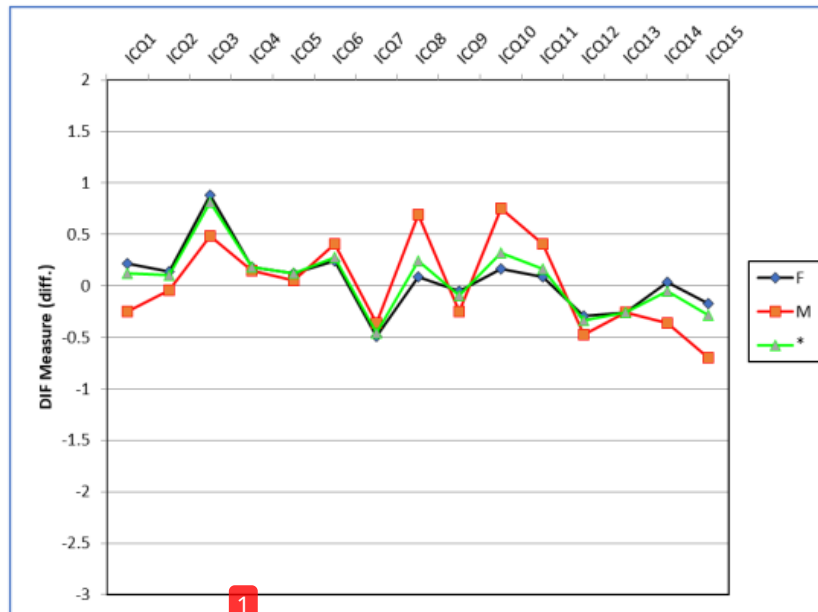
a. Learner's Perspective



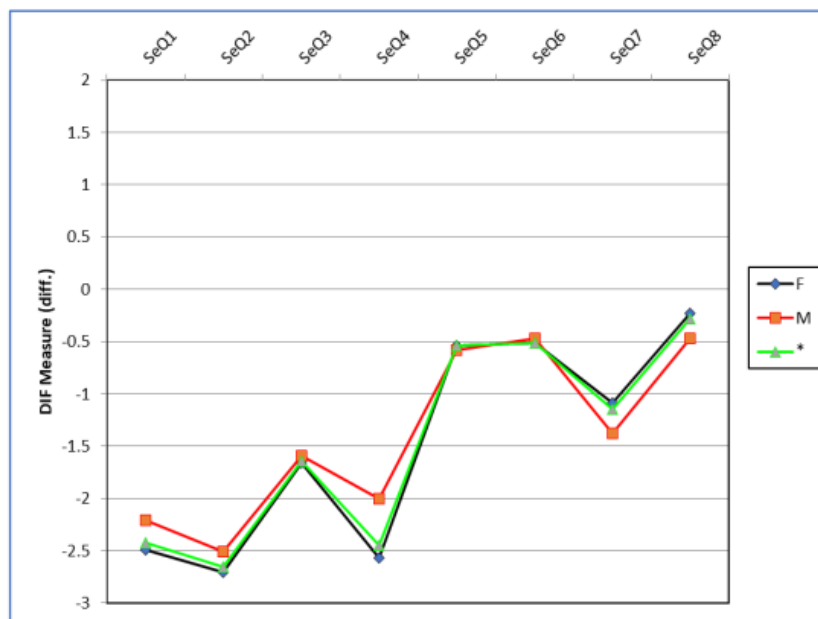
b. Instructor Attitudes



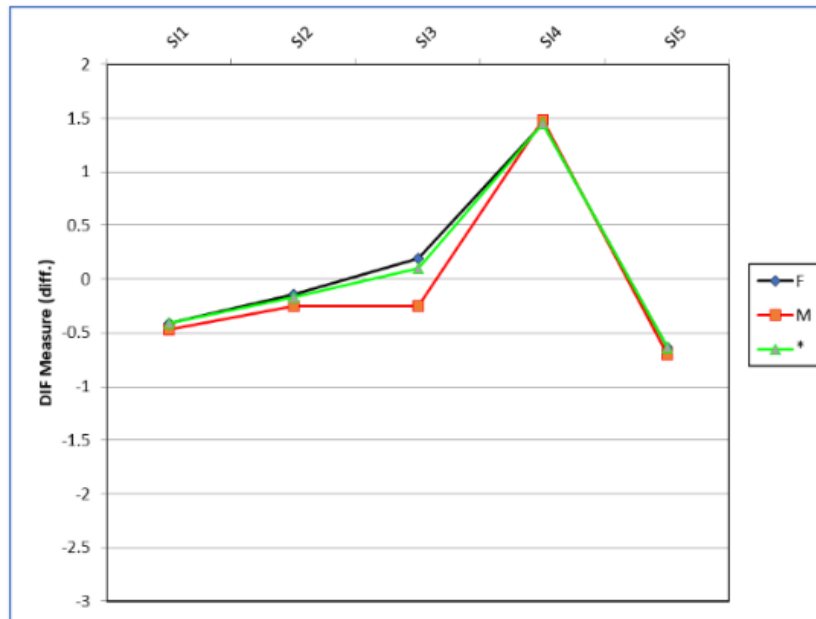
c. System Quality



1
d. Information Content Quality



e. Service Quality



f. Supportive Issues

Note(s): F = Perempuan, M = Laki-laki

Gambar 11. DIF Plot Item di HELAM Berdasarkan Gender

Berdasarkan Tabel 11, lima belas item terindikasi memiliki DIF pada Gender. Empat item (LP1, LP3, LP7 dan LP10) ditemukan memiliki DIF pada dimensi Perspektif Pembelajar. Empat item (IA1, IA2, IA6 dan IA11) pada dimensi Sikap Instruktur terindikasi terinfeksi DIF. Sedangkan pada dimensi System Quality, item SyQ13 dan SyQ15 mengalami DIF. Pada dimensi Kualitas Isi Informasi, item ICQ2, ICQ8, ICQ10, ICQ15 menunjukkan adanya DIF. Item SeQ4 pada dimensi Kualitas Pelayanan juga terindikasi memiliki DIF.

Diskusi

Sifat psikometrik skala HELAM dianalisis menggunakan perangkat lunak Winstep 4.6.1 [44]. Setelah disaring, diketahui bahwa 75% responden digunakan dalam analisis karena memiliki respon yang sesuai.

Selain menyaring orang, penyaringan dilakukan pada skala penilaian Likert 4 poin yang digunakan. Sifat psikometrik pertama yang dilaporkan adalah fungsionalitas skala penilaian Likert. Analisis skala penilaian ini sangat penting untuk melihat apakah 4 penilaian yang digunakan pada skala HELAM versi adaptasi dapat dipahami dengan baik oleh responden [45], [46]. Hasil analisis skala penilaian yang dilakukan menunjukkan bahwa skala penilaian Likert 4 poin yang digunakan pada HELAM dalam konteks Indonesia kurang berfungsi. Tidak ada satupun responden yang memilih “Tidak Setuju”. Hal ini menunjukkan bahwa pada implementasi HELAM selanjutnya, penilaian skala Likert yang direkomendasikan adalah 3 poin. Fungsionalitas skala penilaian diperkuat secara visual pada Gambar 8. Ketiga penilaian yang ditunjukkan pada Gambar 8 menunjukkan bahwa probabilitas 3 penilaian memiliki nilai probabilitas respons lebih dari 0,5 dan menunjukkan bahwa responden memahami. Hal ini belum dilaporkan dalam versi asli yang dikembangkan oleh Ozkan dan Koseler [26], [29].

Sifat psikometrik yang kedua adalah unidimensionalitas. Unidimensionalitas digunakan untuk melihat apakah HELAM mengukur suatu kemampuan atau tidak [47], [48]. Unidimensi ditentukan dengan menggunakan Analisis Komponen Utama dari residu [49]. Skor varians mentah (*raw variance score*) mendekati nilai yang diharapkan dalam model Rasch sebesar 34,0%. Nilai *Unexplained variance* pada kontras ke-1 sebesar 4,0% setara dengan nilai Eigen sebesar 3,78, yang lebih besar dari kriteria 2 [40]. Hal ini menunjukkan tidak adanya *noise* pengukuran, namun terdapat item yang berasal dari variabel yang berbeda. Hasil penyelidikan lebih lanjut membuktikan bahwa item-item yang dicurigai bukan berasal dari variabel lain karena nilai *loading* seluruh item tidak

melebihi 0,60. Jadi, dapat disimpulkan bahwa skala HELAM yang diadaptasi mengukur satu kemampuan.

Selain itu, penilaian reliabilitas skala HELAM memberikan hasil yang sangat menggembirakan. Statistik ringkasan digunakan untuk mengevaluasi keandalannya, dengan temuan yang menunjukkan keandalan yang sangat baik [50]. Nilai keandalan orang sebesar 0,96 menunjukkan konsistensi yang luar biasa di antara responden dalam tanggapan mereka terhadap setiap item dalam HELAM, yang menggarisbawahi keandalan umpan balik mereka. Secara bersamaan, skor reliabilitas item sebesar 0,96 menandakan kualitas item yang tinggi dalam HELAM, memperkuat reliabilitas skala secara keseluruhan [51]. Konsistensi internal yang kuat ini diperkuat oleh nilai *Cronbach's Alpha* sebesar 0,97, yang menandakan keselarasan yang luar biasa antara responden dan item [46]. Hal ini memperkuat konsistensi skala HELAM, selaras dengan temuan yang dilaporkan oleh Ozkan dan Koseler ketika menggunakan pendekatan teori tes klasik [29].

Nilai indeks pemisahan orang (*Person separation*) melebihi 2 [44] menunjukkan sensitivitas skala HELAM yang luar biasa dalam membedakan kemampuan responden. HELAM dengan baik dapat mengklasifikasikan kemampuan responden ke dalam empat tingkat berbeda. Secara bersamaan, indeks pemisahan item menunjukkan bahwa item dalam HELAM dapat dikategorikan ke dalam lima tingkat yang berbeda. Rentang keragaman lima tingkat yang luas ini secara signifikan berkontribusi pada pemahaman komprehensif tentang kesulitan soal, mulai dari sangat mudah hingga sangat sulit, dengan rentang minimal tiga tingkat diperlukan untuk memberikan kesempatan yang adil kepada responden untuk menjawab dengan benar [52], [53].

Analisis sebaran orang dan item pada peta *Wright* dilakukan untuk mengevaluasi validitas konstruk dan sensitivitas skala HELAM. Peta *Wright* digunakan untuk memetakan tingkat kemampuan item dan kemampuan responden secara hierarkis dalam skala logit yang sama [54]. Soal yang tingkat kesulitannya paling rendah terletak di kanan bawah, dan soal yang tingkat kesulitannya paling tinggi ada di kanan atas. Orang yang berkemampuan rendah ditempatkan pada sisi kiri bawah dan orang yang berkemampuan tinggi ditempatkan pada sisi kiri atas. Penggaris logit membentang dari -3 logit hingga 5 logit. Berdasarkan Gambar 9, sebaran tingkat kesukaran soal cukup baik karena soal tersebar merata dari yang mudah hingga yang tersulit. Namun ada beberapa item yang perlu direvisi. Misalnya butir soal SeQ1, SeQ2, SeQ3, dan SeQ4 perlu dimodifikasi untuk meningkatkan tingkat kesulitan butir soal tersebut agar lebih fungsional dalam menjaring informasi sesuai tingkat kemampuan guru. Selain itu, soal-soal yang mempunyai tingkat kesulitan yang sama dapat dimodifikasi agar mempunyai tingkat kesulitan yang lebih tinggi agar lebih berfungsi dalam menjaring informasi dari guru yang berkemampuan tinggi. Pemerataan item pada setiap tingkat kesulitan item akan meningkatkan sensitivitas dan reliabilitas instrumen [55] serta mendukung validitas konstruk.

Sifat psikometrik selanjutnya adalah *item fit*. Item yang sesuai akan memberikan kontribusi yang signifikan terhadap definisi konstruk secara umum [56]. Kesesuaian item dievaluasi berdasarkan kriteria *outfit MNSQ*, *outfit ZSTD* dan *Pt. Mea Corr*. Nilai *MNSQ* yang ideal adalah 1 dan nilai 0,50 – 1,50 merupakan rentang yang baik menunjukkan ukuran yang produktif [44], [51], [57]. Dari 65 item, 19 item teridentifikasi tidak sesuai dengan model Rasch karena nilai *MNSQ* di luar rentang. Sembilan belas item perlu direvisi untuk penelitian selanjutnya karena nilai *Pt. Mea Corr*.

menunjukkan nilai positif meskipun berada di luar rentang penerimaan, 0,40–0,80 [46], [51]. *Pt. Mea Corr.* pada rentang penerimaan menunjukkan bahwa semua item berfungsi dalam arah yang sama untuk memprediksi sifat laten skala HELAM [58]. Hal ini berkontribusi pada tingkat validitas konstruk yang baik [32].

Skala yang baik harus memberikan probabilitas yang sama dalam menjawab benar antar kelompok subjek. Kemungkinan merespons suatu pernyataan tidak boleh dipengaruhi oleh atribut subjek, seperti gender [59]. Skala yang memberikan kemungkinan jawaban berbeda terhadap subjek akan menimbulkan bias. Item tersebut akan berfungsi berbeda, dan informasi yang diperoleh tidak akurat. Bias pengukuran akan menyebabkan tes menjadi invarian atau tidak setara antar kelompok [60]. Bias akan berdampak pada menurunnya validitas skala pengukuran. Bias item di HELAM dievaluasi melalui *Differential Item Functioning* (DIF).

Analisis DIF dilakukan untuk melihat tren item dalam HELAM berdasarkan atribut gender. Butir soal bebas bias jika nilai kontras DIF (nilai penerimaan antara -0,5 hingga 0,5), *Rasch-Welch t* (nilai penerimaan antara -2 hingga 2), dan Probabilitas *Mantel Haenszel* (nilai penerimaan > 0,05) berada dalam rentang penilaian.

Suatu item dipengaruhi oleh DIF jika ketiga indeks berada di luar rentang penerimaan secara bersamaan. Dari lima belas item yang terindikasi tertular DIF, hanya item LP10 yang tertular DIF karena nilai kontras DIF berada di luar rentang $\pm 0,5$, nilai *Rasch-Welch t* berada di luar rentang ± 2 , dan nilai probabilitas berada di luar rentang $< 0,05$. Apabila responden laki-laki dan perempuan dengan tingkat kemampuan yang sama memberikan tanggapan yang berbeda, maka fungsi item tersebut berbeda [61].

Gambar 11a memvisualisasikan bahwa item LP10 mengalami bias gender, sedangkan Gambar 11b–11f tidak menunjukkan adanya bias gender. Berdasarkan Gambar 11a, dapat dilakukan evaluasi bias item lebih lanjut. Pernyataan pada butir LP10 cenderung lebih berpihak pada laki-laki dibandingkan perempuan. Sedangkan keadaan ini berbanding terbalik dengan 9 item lainnya pada dimensi Perspektif Pembelajar. Oleh karena itu, butir soal yang terjangkit DIF perlu dibuang atau dipertahankan dengan melakukan revisi. Artinya, 64 item yang diadaptasi dalam konteks budaya Indonesia dapat dinilai dengan cara yang sama untuk responden laki-laki dan perempuan [61]. Informasi ini melengkapi sifat psikometrik yang belum dilaporkan dalam penelitian sebelumnya oleh Ozkan dan Koseler [29].

D. Kesimpulan

Model Rasch adalah bagian dari teori tes modern yang dapat digunakan untuk mengevaluasi sifat psikometri skala pengukuran. Para ahli bahasa telah merekomendasikan penggunaan skala HELAM secara empiris dalam konteks Indonesia. Berdasarkan evaluasi sifat psikometri, skala HELAM mempunyai ⁶⁵ kualitas yang baik dari segi reliabilitas dan validitas. Sifat psikometrik skala HELAM yang dilaporkan telah mendukung penggunaannya dalam konteks budaya yang berbeda di masa depan. Namun, beberapa item menunjukkan ketidaksesuaian dan dipengaruhi oleh DIF berdasarkan gender. Oleh karena itu, mereka perlu dikeluarkan dari skala atau direvisi. Secara empiris, skala penilaian Likert 3 poin sesuai dengan konteks budaya Indonesia.

E. Potensi penelitian mendatang

Penelitian ini mempunyai kontribusi yang signifikan terhadap informasi sifat psikometri HELAM dan memberikan informasi baru mengenai sifat psikometri HELAM dalam konteks budaya lain, khususnya dalam konteks budaya Indonesia. Namun, karena pengumpulan data baru tidak mungkin dilakukan dalam penelitian ini, maka diperlukan penelitian lebih lanjut untuk merevisi dan mengevaluasi sifat psikometrik item yang tidak sesuai, terpengaruh oleh DIF, dan meningkatkan tingkat kesulitan item untuk mengisi kesenjangan lokasi item di tingkat yang lebih tinggi. kesulitan. Direkomendasikan agar fungsi skala penilaian Likert 4 poin yang digunakan dalam konteks budaya lain harus dievaluasi. Selanjutnya, modifikasi penggunaan skala penilaian Likert yang lebih dari 4 poin harus dipertimbangkan.

F. Daftar Pustaka

- [1] M. Daumiller, R. Rinas, J. Hein, S. Janke, O. Dickhäuser, and M. Dresel, 'Shifting from face-to-face to online teaching during COVID-19: The role of university faculty achievement goals for attitudes towards this sudden change, and their relevance for burnout/engagement and student evaluations of teaching quality', *Comput. Hum. Behav.*, vol. 118, p. 106677, 2021, doi: <https://doi.org/10.1016/j.chb.2020.106677>.
- [2] A. El-Soussi, 'The shift from face-to-face to online teaching due to COVID-19: Its impact on higher education faculty's professional identity', *Int. J. Educ. Res. Open*, vol. 3, p. 100139, 2022.
- [3] A. E. Hallin, H. Danielsson, T. Nordström, and L. Fälth, 'No learning loss in Sweden during the pandemic: Evidence from primary school reading assessments', *Int. J. Educ. Res.*, vol. 114, p. 102011, 2022, doi: <https://doi.org/10.1016/j.ijer.2022.102011>.
- [4] B. Sezer and R. Yilmaz, 'Learning management system acceptance scale (LMSAS): A validity and reliability study', *Australas. J. Educ. Technol.*, vol. 35, no. 3, pp. 15–30, 2019.
- [5] J. Li and W. Che, "Challenges and coping strategies of online learning for college students in the context of COVID-19: A survey of Chinese universities," *Sustain. Cities Soc.*, vol. 83, p. 103958, Aug. 2022, doi: [10.1016/j.scs.2022.103958](https://doi.org/10.1016/j.scs.2022.103958).

BAB 4

Validasi Instrumen Motivasi Belajar Fisika I-PMQ

A. Perumusan Masalah

Mengembangkan tes kognitif dan non-kognitif merupakan hal yang rumit, sehingga memerlukan tinjauan literatur yang ekstensif dan memerlukan waktu yang lama [1]. Tugas ini menjadi semakin menantang seiring dengan meningkatnya permintaan akan instrumen sensitif budaya di berbagai bidang. Instrumen yang dirancang untuk satu konteks budaya seringkali memerlukan adaptasi bagi konteks budaya lain untuk memastikan relevansi dan kemanjurannya. Tren yang berkembang dalam mengadaptasi instrumen-instrumen ini di berbagai disiplin ilmu dan konteks budaya dibuktikan oleh penelitian terbaru (misalnya, [2]–[6]).

Namun, adaptasi melibatkan lebih dari sekedar penerjemahan; hal ini memerlukan pertimbangan yang cermat terhadap nuansa sosio-kultural dari khalayak sasaran. Prosedur adaptasi yang ditetapkan, seperti yang diusulkan oleh Hambleton Hambleton [7] dan Lee *et al.* [8], menawarkan pendekatan sistematis, namun juga mengungkap kompleksitas proses ini. Permulaan pandemi ini telah mempercepat peralihan cara pendidikan, beralih secara tiba-tiba dari pembelajaran tatap muka tradisional ke pembelajaran yang sepenuhnya daring. Perubahan mendadak ini

berdampak pada pendidik dan siswa, mengubah ruang belajar fisik dan psikologis dan akibatnya mempengaruhi motivasi siswa [9]. Penelitian menunjukkan bahwa siswa yang termotivasi lebih efektif dalam konteks belajar mandiri [10], [11]. Oleh karena itu, menilai dan memahami motivasi belajar menjadi penting, khususnya dalam keadaan yang berubah ini.

Science Motivation Questionnaire (SMQ) oleh Glynn *et al.* [12], alat yang terbukti efektif dalam konteks sains dan non-sains, telah mengalami revisi signifikan dari format aslinya yang terdiri dari 30 item, lima dimensi menjadi SMQII yang lebih sederhana dengan 25 item dalam lima faktor [12], [13]. Fleksibilitas instrumen ini ditunjukkan melalui adaptasinya di lingkungan budaya yang beragam, termasuk Yunani [14], Turki [15], Spanyol [16], Republik Ceko [17], dan Cina [18]. Di Indonesia, adaptasi khusus telah dilakukan untuk mahasiswa kedokteran gigi oleh Wardhany *et al.* [19] dan untuk mahasiswa sains oleh Aini *et al.* [20]. Namun, meski informatif, adaptasi ini belum secara komprehensif mengatasi sifat psikometrik SMQII, khususnya dalam konteks model Rasch dan fungsionalitas skala Likert.

Selain itu, adaptasi SMQII dalam disiplin akademik tertentu seperti Fisika masih terbatas, tidak seperti penerapannya dalam Matematika [10], Kimia [14], [15], dan Biologi [17]. Misalnya, Fiorella *et al.* [10] mengubah SMQII menjadi Kuesioner Motivasi Matematika (*Mathematics Motivation Questionnaire*, MMQ), sementara Salta dan Koulougliotis [14] serta Cetin-Dindar dan Geban [15] mengembangkan versi yang berfokus pada kimia. Janštová dan Šorgo [17] menyederhanakan SMQ asli menjadi Kuesioner Motivasi Biologi (*Biology Motivation Questionnaire*, BMQ) yang terdiri dari 14 item yang lebih ringkas. Berdasarkan adaptasi ini, penelitian kami berfokus pada penyesuaian BMQ untuk bidang Fisika.

Mengingat manfaat praktis dari kuesioner yang lebih pendek, seperti tingkat partisipasi responden yang lebih tinggi [17], kami bertujuan untuk menyeimbangkan keringkasan dan kelengkapan. Meskipun teori tes klasik, yang digunakan oleh Janštová dan Šorgo [17], memberikan beberapa wawasan, teori ini gagal dalam melaporkan properti psikometrik terperinci, seperti keandalan tingkat item dan fungsionalitas skala Likert. Untuk mengatasi kesenjangan ini, penelitian kami menggunakan model Rasch untuk memberikan pemahaman yang lebih bernuansa tentang kualitas psikometrik kuesioner motivasi dalam Fisika, sehingga membangun dan memperluas literatur yang ada.

B. Metodologi

Penelitian survei-kuantitatif ini mematuhi pedoman model Rasch, yang menunjukkan bahwa ukuran sampel antara 50 dan 250 peserta sesuai untuk mencapai model yang sesuai [21]. Untuk memastikan stabilitas data dengan akurasi 0,5 logit pada tingkat kepercayaan 95%, penelitian kami merekrut 107 siswa SMA dan SMK Negeri, dengan sebaran gender 66 laki-laki (61,7%) dan 41 perempuan (38,3%). Kami mengecualikan dua siswa (1,8%) dengan skor ekstrem untuk menjaga validitas statistik temuan kami [22], [23].

PMQ diadaptasi dengan cermat dari BMQ [17], menjalani proses *forward-backward translation* yang ketat oleh dosen bahasa Inggris berpengalaman yang berspesialisasi dalam linguistik terapan [6]. Terminologi biologi secara hati-hati dikontekstualisasikan ke dalam domain fisika, memastikan relevansi konten untuk target demografi. Keterbacaan instrumen yang diadaptasi kemudian divalidasi melalui uji coba dengan sekelompok siswa. Skala motivasi PMQ terdiri dari 14 item dalam tiga dimensi—Motivasi (Motivation, Mo), Efikasi Diri (Self-efficacy,

SE), dan Tanggung Jawab (Responsibility, Re)—menggunakan skala Likert 5 poin untuk menangkap respons.

Untuk pengumpulan data, kami menggunakan Google Formulir, yang memfasilitasi proses survei *online* yang efisien dan mudah diakses [24]. Kolaborasi dengan guru sekolah setempat memastikan kelancaran distribusi kuesioner, dan kami menekankan kepada siswa tentang kerahasiaan dan sifat akademis yang tidak memiliki konsekuensi dari partisipasi mereka, sehingga menjamin tanggapan yang sukarela dan tidak memihak. Siswa biasanya menyelesaikan kuesioner dalam waktu 15 hingga 20 menit.

Kualitas psikometrik PMQ diteliti menggunakan model Rasch, dengan data mentah disiapkan di Ms. Excel dan dianalisis melalui perangkat lunak Winsteps [25]. Proses ini mengevaluasi validitas konstruk PMQ, dengan fokus pada relevansi konten, keterwakilan, dan kualitas teknis [26]. Selain itu, kami menyelidiki penggunaan fungsional skala Likert, sebuah faktor penting dalam interpretasi hasil PMQ.

C. Hasil dan Pembahasan

Relevansi konten

Relevansi konten sangat penting dalam memahami hubungan antara soal tes dan konstruk yang diukur, khususnya motivasi siswa dalam belajar fisika. Untuk mengukur hal ini secara akurat, Kuesioner Motivasi Fisika (PMQ) diadaptasi dari Kuesioner Motivasi Biologi (BMQ) yang sudah mapan, sebuah proses yang dilakukan dengan cermat oleh Janštová dan Šorgo [17].

Adaptasi ini melibatkan penyesuaian konten BMQ untuk mencerminkan aspek unik fisika, memastikan bahwa kuesioner sesuai

dengan minat dan tantangan spesifik yang dihadapi siswa fisika. BMQ, yang terkenal dengan konstruksi kokohnya, menjalani proses validasi yang ketat menggunakan pendekatan analisis faktor.

Dari 30 item awal, 14 item diidentifikasi sesuai dengan kriteria model, menunjukkan representasi kuat PMQ terhadap domain konstruk dan memperkuat relevansi kontennya. Adaptasi yang berhasil ini menegaskan kegunaan PMQ dalam menilai motivasi siswa dalam fisika dan menyoroti pentingnya modifikasi instrumen yang tepat dalam penelitian pendidikan. Dengan menangkap esensi keterlibatan siswa dalam fisika, PMQ adalah alat yang berharga dalam memahami dan meningkatkan pengalaman belajar di bidang ini.

Keterwakilan

Indeks strata (H) dan pemisahan atau *Separation* (G) digunakan untuk menilai keterwakilan Kuesioner Motivasi Fisika (PMQ). Kedua indeks ini ditampilkan pada Tabel 12.

Gambar 12. Ringkasan statistik PMQ

	Measure		Separation	Reliability	Cronbach's α
	Mean	SD			
Item	0.00	0.66	3.61	0.93	0.92
Person	2.33	1.60	2.82	0.89	

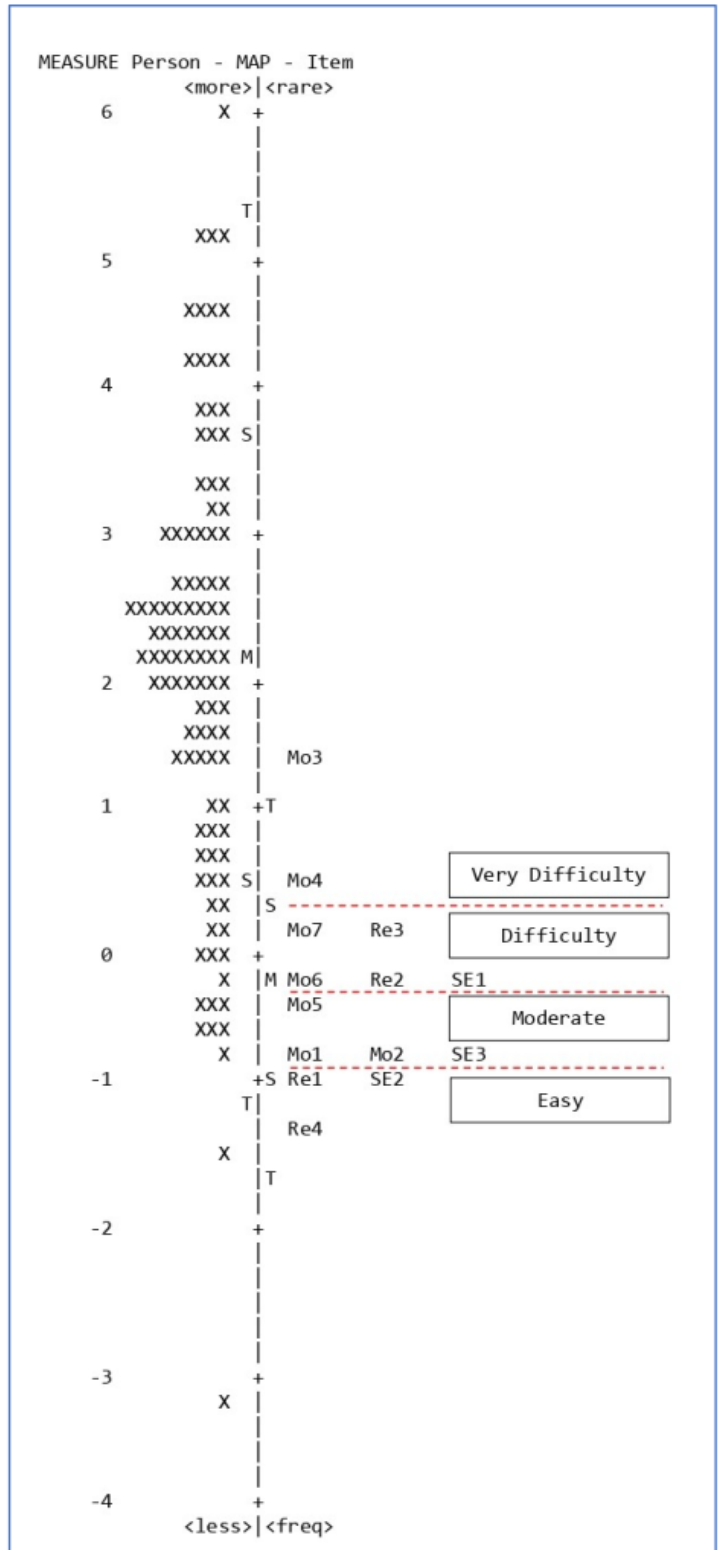
Pada Tabel 12, dua indeks kritis, Strata (H) dan *Separation* (G), berperan penting dalam menilai keterwakilan Kuesioner Motivasi Fisika (PMQ). Indeks-indeks ini menunjukkan hasil yang menjanjikan, yang dirancang khusus untuk mengevaluasi kemampuan instrumen dalam membedakan berbagai tingkat kesulitan soal dan kemampuan responden.

Indeks H dan G, dengan nilai masing-masing 5,14/4,09 dan 3,61/2,82, tidak hanya melampaui nilai batas yang ditetapkan [27], [28]

namun juga menggarisbawahi kemampuan penilaian PMQ yang berbeda-beda. Indeks H yang dihitung menggunakan rumus $H=(4 \times G+1)/3$ [29], semakin memperkuat hal tersebut. Terkait reliabilitas, PMQ menunjukkan ketahanan, sedangkan reliabilitas orang dan soal menunjukkan tingkat konsistensi yang tinggi dalam jawaban responden dan kualitas soal tes. Temuan ini dilengkapi dengan tingkat kepercayaan masing-masing sebesar 89% dan 93% untuk *person* dan *item measure*, serta skor alpha Cronbach yang mengesankan sebesar 0,92 [30], yang menyoroti kualitas interaksi yang sangat baik antara item dan orang.

Analisis lebih lanjut mengenai keterwakilan konten PMQ melibatkan pemeriksaan kuantitas item, hierarki distribusi, dan kesenjangan dalam tingkat kesulitan item [26]. Meskipun Gambar 13 menunjukkan jumlah item yang memadai dan terdistribusi dengan tepat di berbagai tingkat kesulitan,

Gambar 13 juga menyoroti kesenjangan yang mencolok antara item Mo3 dan Mo4 (0,72 logit). Perbedaan ini memerlukan penyelidikan lebih lanjut untuk memahami dampaknya terhadap efektivitas PMQ secara keseluruhan. Namun demikian, temuan ini secara kolektif menegaskan keterwakilan konten PMQ yang kuat, membuka jalan bagi penggunaannya yang andal dalam konteks pendidikan untuk mengukur motivasi siswa dalam fisika.



Gambar 13. Peta Wright motivasi belajar siswa

Kualitas Teknis

Kesesuaian data empiris dengan model Rasch, khususnya melalui nilai *infit* dan *outfit mean-square (MnSq)*, merupakan aspek penting dalam menilai kualitas teknis Kuesioner Motivasi Fisika (PMQ). Meskipun tidak ada standar yang diterima secara universal untuk kesesuaian statistik dalam pengukuran Rasch [31], nilai MnSq yang mendekati 1 umumnya dianggap ideal, dengan kisaran 0,5 hingga 1,5 sering disebut sebagai indikasi kesesuaian yang baik [26].

Berdasarkan data yang disajikan pada Tabel 12, PMQ secara umum berada pada rentang yang dapat diterima, dengan pengecualian pada item Mo3, yang menunjukkan nilai MnSq sedikit lebih tinggi, yaitu di atas 1,5 namun di bawah 2. Penyimpangan ini menunjukkan bahwa responden kurang tanggap terhadap item Mo3 karena sensitivitas atau tingkat kesulitan item. Menariknya, meskipun nilai MnSq tinggi, Model S.E. di bawah 0,30, menunjukkan pemahaman responden yang kuat terhadap item tersebut. Kontras ini menyoroti sifat yang berbeda dalam menilai validitas kuesioner dan menunjukkan bahwa meskipun Mo3 menyimpang dari kesesuaian ideal, dampaknya terhadap pemahaman secara keseluruhan mungkin minimal.

Singkatnya, meskipun ada sedikit penyimpangan yang diamati pada item Mo3, PMQ secara umum sangat mematuhi kriteria kualitas teknis yang ditetapkan oleh model Rasch. Hal ini menunjukkan bahwa PMQ adalah alat canggih yang mampu mengukur motivasi siswa dalam skala secara efektif dengan presisi teknis yang tinggi. Oleh karena itu, analisis keseluruhan mendukung kegunaan PMQ dalam penilaian pendidikan, dengan potensi perbaikan kecil untuk lebih meningkatkan efektivitasnya.

Tabel 12. Model kecocokan PMQ 14-item

Item	Measure (Logit)	Model S.E (Logit)	Infit		Outfit	
			MnSq	Zstd	MnSq	Zstd
Mo1	-0.42	0.17	1.04	0.33	1.15	1.00
Mo2	-0.48	0.17	1.08	0.65	1.06	0.44
Mo3	1.59	0.15	1.56	3.44	1.62	3.65
Mo4	0.87	0.16	0.65	-2.75	0.64	-2.81
Mo5	-0.15	0.17	0.86	-1.04	0.84	-1.10
Mo6	0.14	0.17	1.21	1.46	1.16	1.11
Mo7	0.44	0.17	1.21	1.43	1.15	1.05
SE1	0.11	0.17	1.06	0.46	1.04	0.34
SE2	-0.54	0.18	1.07	0.55	1.11	0.76
SE3	-0.42	0.17	0.87	-0.96	0.85	-1.00
Re1	-0.70	0.18	0.79	-1.63	0.73	-1.82
Re2	0.14	0.17	0.80	-1.44	0.83	-1.23
Re3	0.44	0.17	0.91	-0.62	0.91	-0.60
Re4	-1.02	0.18	0.74	-2.05	0.69	-1.93

Aspek Struktural Validitas Konstruk

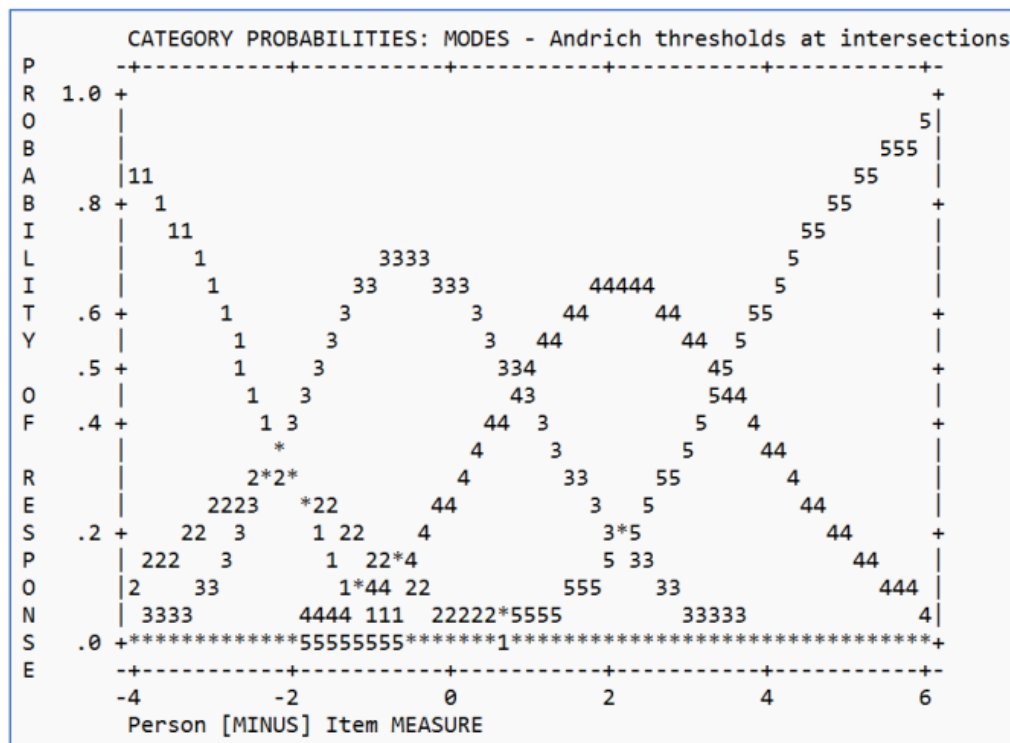
Unidimensionalitas, sebuah aspek penting dalam validitas tes, menyiratkan bahwa penilaian hanya mengukur satu sifat dasar [32]–[35]. Konsep ini sangat penting dalam mengevaluasi aspek 83ingkat83al validitas konstruk PMQ [26]. Untuk menyelidiki unidimensi PMQ, Analisis Komponen Utama Residual (PCAR) digunakan [36], [37]. Analisis mengungkapkan bahwa *raw variance* yang dijelaskan oleh pengukuran dalam PMQ adalah 51,7%, melampaui ambang batas yang direkomendasikan Rasch sebesar 40% [38]. Persentase yang 83ingki ini menunjukkan bahwa sebagian besar varians dalam data dapat dikaitkan dengan sifat utama laten yang diukur dengan PMQ, sehingga menegaskan fokusnya pada satu konstruksi. Selain itu, *unexplained variance in the first contrast* mencapai 7,5%, jauh di bawah batas 15%, menunjukkan pengaruh minimal dari dimensi atau sifat lain.

Temuan ini menggarisbawahi unidimensi PMQ, memperkuat validitasnya sebagai ukuran yang secara akurat menargetkan satu sifat

laten tertentu – motivasi siswa dalam fisika. Pengukuran terfokus seperti ini sangat penting untuk memastikan bahwa PMQ memberikan hasil yang bermakna dan dapat ditafsirkan, terutama dalam konteks pendidikan di mana penilaian yang tepat terhadap motivasi siswa sangatlah penting. Namun, penting untuk diketahui bahwa meskipun PCAR merupakan metode yang kuat untuk menilai unidimensionalitas, evaluasi berkelanjutan dan potensi penyempurnaan terhadap PMQ harus dipertimbangkan untuk menjaga relevansi dan keakuratannya dalam mengembangkan lanskap pendidikan.

Fungsi Skala Penilaian Likert

Untuk melengkapi penilaian aspek isi validitas konstruk dalam PMQ, fungsi skala penilaian Likert 5 poin diteliti dengan cermat. Langkah ini penting untuk memastikan bahwa lima opsi pemeringkatan berbeda dan dapat dimengerti oleh responden. Analisis tersebut, yang didukung oleh inspeksi visual pada kurva probabilitas respons pada Gambar 14, menunjukkan adanya kekurangan yang mencolok dalam pemahaman responden terhadap peringkat kedua, 'Tidak Setuju'. Kurva tersebut tidak mencapai puncak yang jelas untuk peringkat ini, sehingga menunjukkan bahwa responden mungkin perlu membedakannya dengan jelas dari peringkat yang berdekatan. Kesimpulan ini secara kuantitatif didukung oleh *Andrich Threshold*, yang mengukur perbedaan antara kategori pemeringkatan yang berdekatan. Ambang batas antara skala Likert 2 dan Likert 3 hanya 0,38 logit, kurang dari 1,4 logit yang biasanya menunjukkan perbedaan yang jelas. Kurva fungsional, sesuai konvensi [35], harus menampilkan puncak pada setiap 84 tingkat peringkat, yang menandakan skala yang dipahami dengan baik dan dimanfaatkan secara efektif.



Gambar 14. Kurva probabilitas fungsionalitas skala peringkat Likert di PMQ

Implikasi dari temuan ini signifikan terhadap kegunaan PMQ. Kurangnya peringkat yang jelas dan fungsional pada tingkat 'Tidak Setuju' menunjukkan potensi penggabungan kategori ini dengan kategori lain untuk menangkap nuansa motivasi responden dengan lebih baik. Penyesuaian ini dapat meningkatkan keakuratan PMQ dalam menafsirkan derajat setuju atau tidak setuju yang diungkapkan responden. Penyempurnaan seperti ini akan meningkatkan keandalan tanggapan dan, akibatnya, validitas alat pengukuran secara keseluruhan.

D. Kesimpulan

Mengadaptasi Kuesioner Motivasi Fisika (PMQ) ke dalam konteks budaya Indonesia merupakan upaya substantif, yang berpuncak pada keberhasilan merepresentasikan domain konstruk dalam 14 item yang

diadaptasi dengan cermat dari Kuesioner Motivasi Biologi (BMQ). Sifat psikometrik PMQ dibuktikan dengan indeks strata (H) dan pemisahan atau *separation* (G) yang mendukung di samping metrik keandalan yang kuat, yang menegaskan keterwakilan tes tersebut. Lebih lanjut memperkuat keunggulan teknisnya, nilai-nilai *Infit* dan *Outfit MnSq* PMQ selaras dengan standar ketat model Rasch. Pada saat yang sama, demonstrasi unidimensionalitas membuktikan validitas konstruk strukturalnya.

E. Potensi penelitian mendatang

Temuan yang menarik adalah bahwa dalam lingkungan budaya Indonesia, skala Likert 4 poin lebih cocok dibandingkan skala 5 poin awal, sehingga berpotensi meningkatkan sensitivitas dan akurasi respons instrumen. Adaptasi ini menggarisbawahi penerapan dan relevansi PMQ bagi siswa sekolah menengah atas di Indonesia, dan menawarkan alat yang berharga untuk menilai dan menumbuhkan motivasi dalam bidang fisika. Penelitian ini menjembatani kesenjangan kritis dengan menjelaskan kualitas alat motivasi dalam pendidikan fisika di tingkat sekolah menengah, menggunakan model Rasch untuk analisis yang cermat. Namun, ruang lingkup untuk penelitian lebih lanjut masih luas.

Penelitian di masa depan didorong untuk memperluas penerapan PMQ ke demografi yang lebih luas, termasuk siswa sekolah menengah pertama dan lembaga pendidikan swasta, untuk memvalidasi efektivitasnya dan memperkaya kemampuan adaptasi budayanya. Upaya-upaya tersebut tidak hanya akan memperluas penerapan instrumen ini tetapi juga berkontribusi pada pemahaman yang lebih mendalam tentang motivasi siswa dalam beragam lanskap pendidikan di Indonesia.

Lampiran 1. Instrumen Jr.MAI

Junior Metacognitive Awareness Inventory (Jr.MAI)

Nama : _____ (Boleh dikosongkan)

Sekolah : _____

Umur : _____ tahun

Jenis kelamin : Laki-laki / perempuan

Petunjuk Pengisian:

Setiap butir pernyataan terdiri dari 5 pilihan jawaban. Berilah respon untuk setiap pernyataan yang menurut anda sesuai dengan kondisi anda.

Kriteria:

1 = tidak pernah

2 = Jarang

3 = Kadang-kadang

4 = Sering

5 = Selalu

Pernyataan:

No	Pernyataan	Kondisi saya				
		1	2	3	4	5
1	Saya tahu ketika saya memahami sesuatu.					
2	Ketika saya menyukai suatu topik, saya mempelajarinya lebih mendalam					
3	Informasi-informasi penting dalam suatu topik sangat saya perhatikan.					
4	Saya dapat mengkondisikan diri saya untuk belajar ketika saya butuh.					
5	Saya bisa belajar maksimal ketika sudah mengetahui sesuatu dari topik itu					

No	Pernyataan	Kondisi saya				
		1	2	3	4	5
6	Saya mengerti apa yang diharapkan guru pada saya untuk dipelajari					
7	Cara belajar yang berhasil saya gunakan sebelumnya, akan saya gunakan kembali					
8	Saya menggunakan kekuatan-kekuatan belajar yang saya miliki untuk menutupi kelemahan-kelemahan.					
9	Tanpa melalui proses berpikir, saya kadang-kadang langsung menggunakan strategi-strategi belajar.					
10	Setelah menyelesaikan tugas sekolah, saya bertanya kepada diri sendiri apakah saya sudah mempelajari apa yang sebenarnya mau saya pelajari					
11	Saya memikirkan apa yang sebenarnya ingin saya pelajari dari satu topik itu sebelum saya mulai mempelajarinya.					
12	Ketika saya sedang mempelajari materi yang baru, saya bertanya pada diri sendiri seberapa baik saya melakukannya.					
13	Ketika menyelesaikan tugas sekolah, saya mempertimbangkan beberapa cara penyelesaian lalu memilih cara terbaik					
14	Setelah saya menyelesaikan tugas, saya bertanya pada diri sendiri apakah ada cara lain yang lebih mudah.					
15	Untuk membantu memahami materi ketika belajar, saya membuat gambar-gambar atau diagram-diagram.					
16	Strategi-strategi belajar yang saya gunakan bisa saja berbeda tergantung pada tugasnya.					
17	Sebelum mulai mengerjakan tugas, saya memastikan apa saja yang sudah harus dilakukan					
18	Saya memeriksa kembali pekerjaan saya untuk memastikan bisa selesai tepat waktu					

Lampiran 2. Instrumen POSTOL

Perception of Students Towards Online Learning (POSTOL)

A. Demografi Responden:

Nama : _____ (Boleh dikosongkan)

Sekolah : _____

Kelas : _____

Umur : _____ tahun

Jenis kelamin : Laki-laki / perempuan

Apakah Anda memiliki smartphone?

- a. Ya
- b. Tidak

Apakah Anda memiliki Akun Facebook?

- a. Ya
- b. Tidak

Apakah Anda memiliki Akun Instagram?

- a. Ya
- b. Tidak

Frekuensi penggunaan Internet setiap hari:

- a. Jarang (<1 jam)
- b. Kadang-kadang (1-2 jam)
- c. Biasanya (3-4 jam)
- d. Sering (> 5 jam)

B. Petunjuk Pengisian:

Setiap butir pernyataan terdiri dari 5 pilihan jawaban. Berilah respon untuk setiap pernyataan yang menurut anda sesuai dengan kondisi anda.

Kriteria:

1 = Sangat Tidak Setuju

2 = Tidak Setuju

3 = Ragu-Ragu

4 = Setuju

5 = Sangat Setuju

C. Pernyataan:

No	Pernyataan	Kondisi saya				
		1	2	3	4	5
1	Guru harus ramah dan mudah ditemui.					
2	Guru harus mendorong interaksi siswa					
3	Guru harus menyediakan sumber belajar online yang memadai					
4	Guru harus memecahkan masalah yang muncul secara efisien					
5	Guru merespon dengan cepat saat ada pertanyaan di forum diskusi					
6	Pembelajaran ini membantu saya menggunakan sumber internet dengan lebih efisien					
7	Bagi saya berbagi ilmu melalui diskusi online adalah ide yang bagus					
8	Diskusi online memungkinkan siswa untuk bertukar pikiran dan komentar					
9	Saya mendapatkan manfaat dari penggunaan aplikasi interaktif					
10	Membaca-baca pekerjaan teman sekelas akan membantu meningkatkan kualitas pekerjaan saya sendiri					
11	Saya dapat membedakan antara materi pelajaran yang sulit dan mudah, kemudian mempelajarinya dengan cara yang berbeda					
12	Saya suka terlibat aktif dalam diskusi kelompok					
13	Memahami materi pelajaran sangat penting bagi saya					
14	Pembelajaran online harus memberikan pengalaman belajar yang lebih baik daripada pembelajaran offline					
15	Saya yakin dapat memperoleh nilai yang lebih baik dalam pembelajaran online daripada saat offline					
16	Saya belajar lebih giat dalam pembelajaran online daripada saat offline					

Lampiran 3. Instrumen I-PMQ

Survey Motivasi dalam Belajar Fisika

Bagian 1.

Nama Sekolah : _____

Jenjang Sekolah : SMP/SMA*

Kelas : _____

Umur : _____

Jenis Kelamin : Laki-laki/Perempuan*

Bagian 2.

Berilah tanda centang pada setiap pertanyaan yang menurut anda sesuai

No.	Pertanyaan	Jawaban	
		Ya	Tidak
1	Apakah Anda memiliki smartphone		
2	Apakah Anda memiliki Akun Facebook		
3	Apakah Anda memiliki Akun Instagram		
4	Frekuensi penggunaan Internet setiap hari	Sering (> 5 jam) Biasa (> 3 s.d 4 jam) Kadang-kadang (1 s.d. 2 jam) Jarang (<1 jam)	

Bagian 3.

Pada bagian ini terdapat 14 butir pernyataan tentang motivasi dalam pembelajaran fisika. Setiap pernyataan membutuhkan persetujuan yang terdiri dari 5 pilihan. Adik-adik dapat memilih salah satu diantaranya.

27

1 = Sangat Tidak Setuju

2 = Tidak Setuju

3 = Ragu-ragu

4 = Setuju

5 = Sangat Setuju

No	Pernyataan	Kondisi saya				
		1	2	3	4	5
1	Menurut saya belajar fisika/IPA dapat membantu karir saya di masa mendatang					
2	Fisika/IPA yang saya pelajari berguna dalam kehidupan sehari-hari					
3	Fisika/IPA yang saya pelajari sejalan/cocok dengan cita- cita saya					
4	Fisika/IPA yang saya pelajari memiliki nilai praktis bagi saya					
5	Menurut saya belajar fisika/IPA dapat membantu memperoleh pekerjaan yang baik dimasa akan datang.					
6	Menurut saya belajar fisika/IPA itu menarik					
7	Saya senang belajar fisika/IPA					
8	Saya percaya, saya dapat memperoleh nilai yang bagus dalam pelajaran fisika/IPA					
9	Saya berharap dapat melakukan sesuatu yang lebih baik dari siswa lain saat belajar fisika/IPA.					
10	Saya yakin saya bisa menguasai pengetahuan dan keterampilan dalam mata pembelajaran fisika/IPA					
11	Saya mempersiapkan diri dengan baik untuk penilaian mata pelajaran fisika/IPA					
12	Saya menggunakan strategi belajar untuk memastikan saya belajar fisika/IPA dengan baik					
13	Saya belajar cukup keras dalam pelajaran fisika/IPA					
14	Memperoleh nilai fisika/IPA yang baik merupakan hal yang penting bagi saya.					

Validasi Instrumen Penelitian Menggunakan **PEMODELAN RASCH** di Bidang Sosial



Dalam proses pendidikan, terutama di tingkat menengah dan tinggi, evaluasi yang akurat terhadap kemampuan siswa sangatlah penting. Instrumen yang valid dan reliabel merupakan kunci untuk mendapatkan data yang sah, yang pada akhirnya akan mendukung pengambilan keputusan yang tepat dalam bidang pendidikan. Oleh karena itu, buku ini menyajikan pendekatan pemodelan Rasch sebagai metode yang unggul dalam validasi instrumen penelitian. Pemodelan Rasch bukanlah metode yang baru, namun penerapannya dalam konteks pendidikan di Indonesia masih memerlukan banyak perhatian dan adaptasi.

Buku ini memberikan penjelasan yang komprehensif tentang bagaimana pemodelan Rasch dapat digunakan untuk validasi berbagai instrumen penelitian, mulai dari instrumen kesadaran metakognitif hingga persepsi belajar online. Kami berharap buku ini dapat menjadi referensi bagi para peneliti, akademisi, dan praktisi pendidikan yang ingin meningkatkan kualitas penelitian mereka melalui validasi instrumen yang lebih baik.

Buku ini terdiri dari beberapa bab yang masing-masing membahas validasi instrumen yang berbeda, yaitu Jr.MAI untuk kesadaran metakognitif, POSTOL untuk persepsi belajar online, HELAM untuk penilaian e-learning, dan I-PMQ untuk motivasi belajar fisika. Setiap bab dilengkapi dengan metodologi yang detail, hasil dan pembahasan yang mendalam, serta kesimpulan yang jelas, sehingga pembaca dapat memahami proses validasi instrumen dengan menggunakan pemodelan Rasch secara menyeluruh.



Penerbit :
Pustaka Pranala
Jln. Nogabondo No. 07, Rejowinangun
Kotagede, Yogyakarta 55171
Telp. (0274) 444710
Anggota Ikatan Penerbit Indonesia (IKAPI)



Pemodelan Rasch Bidang Sosial rev2-press copy.pdf

ORIGINALITY REPORT

12%

SIMILARITY INDEX

PRIMARY SOURCES

1	Trikinasih Handayani, Moh Irma Sukarelawan, Mohd Effendi Ewan Mohd Matore, Much Fuad Saifuddin. "Adaptation and evaluation of psychometric properties of HELAM using the Rasch model in the Indonesian context", International Journal of Evaluation and Research in Education (IJERE), 2023 Crossref	289 words — 2%
2	journal.unnes.ac.id Internet	103 words — 1%
3	pdfcoffee.com Internet	96 words — 1%
4	journal1.uad.ac.id Internet	85 words — 1%
5	etheses.uin-malang.ac.id Internet	82 words — < 1%
6	eudl.eu Internet	66 words — < 1%
7	ijcs.stmikindonesia.ac.id Internet	55 words — < 1%
8	repository.penerbiteureka.com Internet	53 words — < 1%
9	es.scribd.com Internet	49 words — < 1%

10	eprints.uad.ac.id Internet	42 words — < 1%
11	insightsociety.org Internet	42 words — < 1%
12	repository.upi.edu Internet	42 words — < 1%
13	Joan Ndung'u, Ilan Vertinsky, Joseph Onyango. "The relationship between social media use, social media types, and job performance amongst faculty in Kenya private universities", Heliyon, 2023 Crossref	41 words — < 1%
14	Muhammad Farrukh Shahzad, Shuo Xu, Weng Marc Lim, Xingbing Yang, Qasim Raza Khan. "Artificial intelligence and social media on academic performance and mental well-being: Student perceptions of positive impact in the age of smart learning", Heliyon, 2024 Crossref	40 words — < 1%
15	Samina Amin, M. Irfan Uddin, Wali Khan Mashwani, Ala Abdulsalam Alarood, Abdulrahman Alzahrani, Ahmed Omar Alzahrani. "Developing a Personalized E-Learning and MOOC Recommender System in IoT-Enabled Smart Education", IEEE Access, 2023 Crossref	35 words — < 1%
16	www.researchgate.net Internet	35 words — < 1%
17	123dok.com Internet	32 words — < 1%
18	www.sci-int.com Internet	32 words — < 1%
19	adoc.pub Internet	

28 words — < 1%

20 eprints.uny.ac.id
Internet

28 words — < 1%

21 id.m.wikipedia.org
Internet

24 words — < 1%

22 repository.usd.ac.id
Internet

24 words — < 1%

23 ejournal.radenintan.ac.id
Internet

21 words — < 1%

24 core.ac.uk
Internet

20 words — < 1%

25 e-journal.uajy.ac.id
Internet

20 words — < 1%

26 ejurnal.umri.ac.id
Internet

18 words — < 1%

27 suhadinet.files.wordpress.com
Internet

18 words — < 1%

28 text-id.123dok.com
Internet

18 words — < 1%

29 eprints.walisongo.ac.id
Internet

17 words — < 1%

30 id.123dok.com
Internet

17 words — < 1%

31 yanti jayanti. "standar isi (kurikulum?", INA-Rxiv,
2019
Publications

17 words — < 1%

-
- 32 Nur Asma Riani Siregar, Susanti, Mariyanti Elvi. "Analisis Model Rasch Disposisi Matematis Mahasiswa pada Program Studi Pendidikan Matematika UMRAH", Jurnal Gantang, 2021
Crossref 16 words — < 1%
-
- 33 docplayer.info
Internet 16 words — < 1%
-
- 34 repository.unj.ac.id
Internet 16 words — < 1%
-
- 35 stikes-nhm.e-journal.id
Internet 16 words — < 1%
-
- 36 wiki.fluidproject.org
Internet 16 words — < 1%
-
- 37 Pacific Rim Objective Measurement Symposium (PROMS) 2014 Conference Proceedings, 2015.
Crossref 14 words — < 1%
-
- 38 digilib.uinkhas.ac.id
Internet 14 words — < 1%
-
- 39 idoc.tips
Internet 13 words — < 1%
-
- 40 Mitha Nurfallah, Trisna Roy Pradipta. "Motivasi Belajar Matematika Siswa Sekolah Menengah Selama Pembelajaran Daring di Masa Pandemi COVID-19", Jurnal Cendekia : Jurnal Pendidikan Matematika, 2021
Crossref 12 words — < 1%
-
- 41 adabaday.multiply.com
Internet 12 words — < 1%
-
- 42 viviovio.blogspot.com
Internet 12 words — < 1%

43	dspace.uui.ac.id Internet	11 words — < 1%
44	repositori.uin-alauddin.ac.id Internet	11 words — < 1%
45	Hanoi Pedagogical University 2 Publications	10 words — < 1%
46	ejournal.iainbukittinggi.ac.id Internet	10 words — < 1%
47	setkab.go.id Internet	10 words — < 1%
48	unpar.ac.id Internet	10 words — < 1%
49	www.moe.gov.my Internet	10 words — < 1%
50	Mitrayani Mitrayani, Saleh Hidayat, Naintyn Novitasari. "PENGARUH MODEL PEMBELAJARAN LEARNING CYCLE 7E TERHADAP HASIL BELAJAR SISWA PADA MATERI KEANEKARAGAMAN HAYATI KELAS X MIA DI SMA NEGERI 10 PALEMBANG", BIOEDUKASI (Jurnal Pendidikan Biologi), 2018 Crossref	9 words — < 1%
51	Shao, Y.. "Tests of soil organic carbon density modeled by InTEC in China's forest ecosystems", Journal of Environmental Management, 200711 Crossref	9 words — < 1%
52	eprints.undip.ac.id Internet	9 words — < 1%
53	id.scribd.com Internet	9 words — < 1%

54	ijphs.iaescore.com Internet	9 words — < 1%
55	repository.ar-raniry.ac.id Internet	9 words — < 1%
56	repository.its.ac.id Internet	9 words — < 1%
57	repository.umy.ac.id Internet	9 words — < 1%
58	staffnew.uny.ac.id Internet	9 words — < 1%
59	wolacom.com Internet	9 words — < 1%
60	Firmansyah Firmansyah. "Motivasi Belajar dan Respon Siswa terhadap Online Learning sebagai Strategi Pembelajaran di Masa Pandemi Covid-19", <i>EDUKATIF : JURNAL ILMU PENDIDIKAN</i> , 2021 Crossref	8 words — < 1%
61	anchor.fm Internet	8 words — < 1%
62	ind.topview-engineering.com Internet	8 words — < 1%
63	jonedu.org Internet	8 words — < 1%
64	logicy.wordpress.com Internet	8 words — < 1%
65	mafiadoc.com Internet	8 words — < 1%
66	pointerfisika.wordpress.com	

	Internet	8 words — < 1%
67	repository.nobel.ac.id Internet	8 words — < 1%
68	repository.uinjkt.ac.id Internet	8 words — < 1%
69	repository.unair.ac.id Internet	8 words — < 1%
70	repository.unwira.ac.id Internet	8 words — < 1%
71	www.jurnal.unsyiah.ac.id Internet	8 words — < 1%
72	www.thieme-connect.com Internet	8 words — < 1%
73	zombiedoc.com Internet	8 words — < 1%
74	Peggy G.H. Kasenda, Grace F. Kaparang, Anthony Pangemanan, Shapely S. Ambalao. "FINANCIAL STRESS, SELF-EFFICACY AND FINANCIAL HELP- SEEKING BEHAVIOUR AMONG COLLEGE STUDENTS DURING PANDEMIC", JMBI UNSRAT (Jurnal Ilmiah Manajemen Bisnis dan Inovasi Universitas Sam Ratulangi)., 2022 Crossref	7 words — < 1%
75	"Psychometric and Structural Evaluation of the Physics Metacognition Inventory Instrument", European Journal of Educational Research, 2020 Crossref	6 words — < 1%
76	Muhammad Ramdhan, Moh. Irma Sukarelawan, M. Anas Thohir, Fitria Arifiyanti. "Junior high school student perception of online learning in pandemic	6 words — < 1%

Covid-19: Gender, social media ownership, and internet access duration perspective", International Journal of Education and Learning, 2022

Crossref

EXCLUDE QUOTES OFF

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES OFF

EXCLUDE MATCHES OFF