

**PENGEMBANGAN INSTRUMEN ASESMEN *HIGHER ORDER THINKING*
SKILLS BERBENTUK *TWO-TIER MULTIPLE CHOICE* PADA MATA PELAJARAN
SEJARAH KEBUDAYAAN ISLAM KELAS XI DI MAN 2 SLEMAN**



Oleh: Siti Nur Khotimah

NIM: 24204011042

TESIS

Diajukan kepada Program Magister (S2)

Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga
untuk Memenuhi Salah Satu Syarat guna Memperoleh Gelar Magister

Pendidikan (M.Pd.)

Program Studi Pendidikan Agama Islam

YOGYAKARTA

2026

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Siti Nur Khotimah
NIM : 24204011042
Jenjang : Magister (S2)
Program studi : Pendidikan Agama Islam

Menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagaian yang dirujuk sumbernya.

Yogyakarta, 19 Mei 2026

Saya yang menyatakan,



Siti Nur Khotimah

NIM. 24204011042

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Siti Nur Khotimah

NIM : 24204011042

Jenjang : Magister (S2)

Program Studi : Pendidikan Agama Islam

Menyatakan bahwa naskah tesis ini secara keseluruhan benarbenar bebas dari plagiasi.
Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 19 Mei 2026

Saya yang menyatakan,



Siti Nur Khotimah

NIM. 24204011042

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

NOTA DINAS PEMBIMBING

Kepada Yth.

Dekan Fakultas Ilmu Tarbiyah dan Keguruan

UIN Sunan Kalijaga

Assalamu'alaikum Warahmatullahi Wabarakatuh

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

PENGEMBANGAN INSTRUMEN ASESMEN *HIGHER ORDER THINKING SKILLS* BERBASIS *TWO-TIER MULTIPLE CHOICE* PADA MATA PELAJARAN SEJARAH KEBUDAYAAN ISLAM KELAS XI DI MAN 2 SLEMAN.

Yang ditulis oleh:

Nama : Siti Nur Khotimah

NIM : 24204011042

Jenjang : Magister (S2)

Program Studi : Pendidikan Agama Islam

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada program Magister (S2) Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta untuk diajukan dalam rangka memperoleh gelar Magister Pendidikan (M.Pd).

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 19 Mei 2026

Pembimbing,



Prof. Dr. H. Tasman Hamami, M.A.

NIP. 19611102 198603 1 003

SURAT PERNYATAAN BERJILBAB

Yang bertanda tangan dibawah ini:

Nama : Siti Nur Khotimah
NIM : 24204011042
Jenjang : Magister (S2)
Program Studi : Pendidikan Agama Islam
Fakultas : Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga

Menyatakan dengan ini, bahwa sesungguhnya saya tidak menuntut kepada program studi Pendidikan Agama Islam Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta (atas pemakaian jilbab dalam ijazah Strata Dua), seandainya suatu hari nanti terdapat instansi yang menolak ijazah tersebut dikarenakan penggunaan jilbab.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya dan penuh kesadaran atas ridha Allah Swt.

Yogyakarta, 19 Mei 2026

Saya yang menyatakan,



10000
METERAI
TEMPEL
E677DAOX001990936

Siti Nur Khotimah

NIM. 24204011042

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PERSETUJUAN TIM PENGUJI

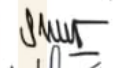
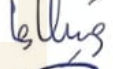
UJIAN TESIS

Tesis Berjudul :

PENGEMBANGAN INSTRUMEN ASESMEN *HIGHER ORDER THINKING SKILLS* BERBENTUK
TWO-TIER MULTIPLE CHOICE PADA MATA PELAJARAN SEJARAH KEBUDAYAAN ISLAM KELAS
XI DI MAN 2 SLEMAN

Nama : Siti Nur Khotimah
NIM : 24204011042
Program Studi : Pendidikan Agama Islam
Fakultas : Ilmu Tarbiyah dan Keguruan

Telah disetujui tim penguji munaqosyah

Ketua/Pembimbing : Prof. Dr. H. Tasman, M.A. ()
Sekretaris/Penguji I : Prof. Dr. H. Abdul Munip, M. Ag. ()
Penguji II : Prof. Dr. H. Sabarudin, M.Si. ()

Diuji di Yogyakarta pada :

Tanggal : 15 Juni 2026
Waktu : 09.30 - 10.45 WIB.
Hasil : A- (94,33)
IPK : 3,90
Predikat : Pujian (Cum Laude)

*coret yang tidak perlu

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS ILMU TARBIYAH DAN KEGURUAN
Jl. Marsda Adisucipto Telp. (0274) 513056 Fax. (0274) 586117 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1822/Un.02/DT/PP.00.9/06/2026

Tugas Akhir dengan judul : PENGEMBANGAN INSTRUMEN ASESMEN *HIGHER ORDER THINKING SKILLS*
BERBENTUK *TWO-TIER MULTIPLE CHOICE* PADA MATA PELAJARAN
SEJARAH KEBUDAYAAN ISLAM KELAS XI DI MAN 2 SLEMAN

yang dipersiapkan dan disusun oleh:

Nama : SITI NUR KHOTIMAH, S.Pd.
Nomor Induk Mahasiswa : 24204011042
Telah diujikan pada : Senin, 15 Juni 2026
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Valid ID: 6a3b5b5ac4d57

Ketua Sidang

Prof. Dr. H. Tasman, M.A.
SIGNED



Valid ID: 6a3cn2b4d7717

Penguji I

Prof. Dr. Abdul Munip, S.Ag., M.Ag.
SIGNED



Valid ID: 6a3a3a3a8e8a1

Penguji II

Prof. Dr. Sabarudin, M.Si
SIGNED



Valid ID: 6a3ccc0bce227

Yogyakarta, 15 Juni 2026

UIN Sunan Kalijaga
Dekan Fakultas Ilmu Tarbiyah dan Keguruan

Prof. Dr. Sigit Purnama, S.Pd.I., M.Pd.
SIGNED

MOTTO

“The mind is not a vessel that needs filling, but wood that needs igniting”
(*Pikiran bukanlah wadah yang perlu diisi, tetapi kayu yang perlu dinyalakan*)

Plutarch¹



¹ Plutarch, “On Listening to Lectures (Trans. Robin Waterfield),” in *Essays by Plutarch* (London: Penguin Classics, 1992), hlm. 18.

PERSEMBAHAN

Dengan iringan doa dan rasa syukur kehadiran Allah Swt., lembar karya ini penulis persembahkan sebagai wujud dedikasi kepada Almamater tercinta, Program Studi Magister Pendidikan Agama Islam, Fakultas Ilmu Tarbiyah dan Keguruan, UIN Sunan Kalijaga Yogyakarta.



ABSTRAK

Siti Nur Khotimah. NIM. 24204011042. *Pengembangan Instrumen Asesmen Higher Order Thinking Skills Berbentuk Two-Tier Multiple Choice pada Mata Pelajaran Sejarah Kebudayaan Islam Kelas XI di MAN 2 Sleman.* Tesis Program Studi Pendidikan Agama Islam (PAI) Program Magister UIN Sunan Kalijaga Yogyakarta, 2026. Pembimbing: Prof. Dr. H. Tasman Hamami, M.A.

Keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skills/HOTS*) merupakan kompetensi esensial yang dituntut dalam pendidikan abad ke-21. Namun, hasil studi internasional PISA 2022 menunjukkan bahwa Indonesia berada di peringkat ke-68 dari 81 negara, mengindikasikan rendahnya kemampuan berpikir kritis peserta didik. Kondisi ini diperparah oleh dominasi instrumen asesmen kognitif tingkat rendah (C1-C2) pada mata pelajaran Sejarah Kebudayaan Islam (SKI). Oleh karena itu, penelitian ini bertujuan untuk: (1) mengembangkan instrumen asesmen *Two-Tier Multiple Choice* (TTMC) berorientasi HOTS pada materi Dinasti Abbasiyah; (2) menguji validitas instrumen berdasarkan *expert judgment*; dan (3) menganalisis kualitas instrumen menggunakan *Model Rasch*; dan (4) mendeskripsikan efektivitas instrumen dalam mengidentifikasi pola keterpahaman konsep (K1-K4) peserta didik.

Metode penelitian yang digunakan adalah *Research and Development* (R&D) dengan model *formative evaluation Tesser* yang meliputi tahap *preliminary*, *self-evaluation*, *prototype*, dan *field test*. Subjek uji coba lapangan terdiri dari 30 peserta didik kelas XI MAN 2 Sleman. Data dianalisis menggunakan *Content Validity Index* (CVI) dan perangkat lunak Jamovi dengan pendekatan *Dichotomous Model* pada *Model Rasch*.

Hasil penelitian menunjukkan bahwa: (1) proses pengembangan instrumen TTMC menghasilkan produk berupa 30 butir soal level kognitif C4 dan C5; (2) instrumen memiliki validitas isi yang baik dengan skor S-CVI/*Average* sebesar 0,88; (3) hasil analisis lapangan menunjukkan kualitas instrumen berdasarkan Model Rasch dengan *person reliability* sebesar 0,629 (cukup) dengan mayoritas butir soal (24 dari 30) dinyatakan *fit* menurut kriteria *Infit/Outfit* MNSQ; (4) instrumen ini secara efektif mampu mengidentifikasi pola keterpahaman konsep peserta didik, dengan kategori paham konsep (K1) sebesar 69,4%, tidak paham (K4) sebesar 13,0%, miskonsepsi (K3) sebesar 9,4%, dan paham parsial (K2) sebesar 8,1%. Instrumen TTMC ini berpotensi layak sebagai alat evaluasi diagnostik untuk mengukur HOTS peserta didik pada materi Dinasti Abbasiyah.

Kata Kunci: Dinasti Abbasiyah, HOTS, *Model Rasch*, Pengembangan Instrumen, Sejarah Kebudayaan Islam, *Two-Tier Multiple Choice*

ABSTRACT

Siti Nur Khotimah. NIM. 24204011042. *Pengembangan Instrumen Asesmen Higher Order Thinking Skills Berbentuk Two-Tier Multiple Choice pada Mata Pelajaran Sejarah Kebudayaan Islam Kelas XI di MAN 2 Sleman.* Thesis of Islamic Religious Education (PAI) Study Program, Master's Program, UIN Sunan Kalijaga Yogyakarta, 2026. Supervisor: Prof. Dr. H. Tasman Hamami, M.A.

Higher Order Thinking Skills (HOTS) are essential competencies demanded in 21st-century education. However, the 2022 PISA results show that Indonesia ranks 68th out of 81 countries, indicating low critical thinking skills among students. This condition is further exacerbated by the dominance of low-level cognitive assessments (C1–C2) in Islamic Cultural History (SKI) subjects. Therefore, this study aims to: (1) develop a HOTS-oriented Two-Tier Multiple Choice (TTMC) assessment instrument for the Abbasid Dynasty material; (2) evaluate the instrument's validity based on expert judgment; (3) analyze its quality using the Rasch Model; and (4) describe the instrument's effectiveness in identifying students' conceptual understanding patterns (K1-K4).

The research adopted a Research and Development (R&D) approach with Tessmer's formative evaluation model, consisting of preliminary, self-evaluation, prototype, and field test stages. The field test was conducted with 30 eleventh-grade students at MAN 2 Sleman. Data were analyzed using the Content Validity Index (CVI) and the Rasch Model via Jamovi software.

The results show that: (1) the development process of the TTMC instrument produced 30 items at the C4 and C5 cognitive levels; (2) the instrument has good content validity with an S-CVI/Average score of 0.88; (3) field test analysis indicates the quality of the instrument based on the Rasch Model, with a person reliability of 0.629 (fair), with the majority of items (24 out of 30) fitting the Infit/Outfit MNSQ criteria; (4) the instrument effectively identifies students' conceptual understanding patterns, with the conceptual understanding category (K1) at 69.4%, lack of understanding (K4) at 13.0%, misconceptions (K3) at 9.4%, and partial understanding (K2) at 8.1%. This TTMC instrument has the potential to be a suitable diagnostic evaluation tool to measure students' HOTS on the Abbasid Dynasty material.

Keywords: Abbasid Dynasty, HOTS, Instrument Development, Islamic Cultural History, Rasch Model, Two-Tier Multiple Choice

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah Swt. yang telah melimpahkan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul “Pengembangan Instrumen Asesmen *Higher Order Thinking Skills* Berbentuk *Two-Tier Multiple Choice* pada Mata Pelajaran Sejarah Kebudayaan Islam Kelas XI di MAN 2 Sleman” ini dengan baik. Shalawat serta salam senantiasa tercurahkan kepada junjungan kita Nabi Muhammad Saw., keluarga, sahabat, dan seluruh umatnya yang istiqomah di jalan-Nya hingga akhir zaman.

Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Pendidikan (M.Pd.) pada Program Studi Pendidikan Agama Islam, Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Penulis menyadari bahwa penyelesaian tesis ini tidak terlepas dari dukungan, bimbingan, dan bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Noorhaidi Hasan, S.Ag., M.A., M.Phil., Ph.D., selaku Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Bapak Prof. Dr. Sigit Purnama, M.Pd., selaku Dekan Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Ibu Dr. Hj. Dwi Ratnasari, M.Ag., selaku Ketua Program Studi Pendidikan Agama Islam Program Magister Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
4. Bapak Prof. Dr. H. Tasman Hamami, M.A., selaku Dosen Pembimbing Tesis yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta motivasi kepada penulis selama proses penyusunan tesis ini dari awal hingga selesai.
5. Dr. Andi Prastowo, S.P.d.I., M.Pd.I., selaku Dosen Penasihat Akademik selama masa perkuliahan penulis.
6. Ibu Dr. Nur Saidah, S.Ag., M.Ag., selaku validator ahli materi yang telah meluangkan waktu untuk memberikan penilaian, masukan, dan saran terhadap instrumen yang dikembangkan, sehingga instrumen ini menjadi lebih baik dan layak digunakan.
7. Ibu Dr. Hj. Endang Sulistyowati, M.Pd.I., selaku validator ahli evaluasi yang telah meluangkan waktu untuk memberikan penilaian, masukan, dan saran terhadap instrumen yang dikembangkan, sehingga instrumen ini menjadi lebih baik dan layak digunakan.

8. Bapak Drs. H. Wiranto Prasetyahadi, M.Pd. selaku Kepala MAN 2 Sleman dan Ibu Hj. Nur Syam'ah, M.Pd. selaku Waka Kurikulum MAN 2 Sleman yang telah memberikan izin kepada penulis untuk melaksanakan penelitian di madrasah yang beliau pimpin.
9. Ibu Hanifah, S. Hum., selaku Guru Mata Pelajaran Sejarah Kebudayaan Islam di MAN 2 Sleman yang telah bersedia menjadi narasumber dan memberikan dukungan penuh selama proses penelitian berlangsung.
10. Staf Tata Usaha Prodi Magister Pendidikan Agama Islam, yang telah membantu memudahkan berbagai urusan administratif berkaitan dengan perkuliahan dan penyusunan tesis ini.
11. Kedua orang tua dan kakak tercinta yang senantiasa memberikan doa, dukungan moral, dan semangat yang tidak pernah padam kepada penulis dalam menyelesaikan studi ini.
12. Seluruh teman-teman Program Studi Pendidikan Agama Islam Program Magister angkatan 2024 yang telah berbagi semangat, pengalaman, dan kebersamaan selama masa perkuliahan.
13. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan kontribusi, baik langsung maupun tidak langsung, dalam penyelesaian tesis ini.

Penulis menyadari sepenuhnya bahwa tesis ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang konstruktif dari berbagai pihak demi perbaikan dan penyempurnaan tesis ini. Penulis berharap semoga tesis ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang evaluasi pembelajaran Pendidikan Agama Islam.

Yogyakarta, 19 Mei 2026

Penulis,



Siti Nur Khotimah

NIM. 24204011042

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN KEASLIAN	ii
PERNYATAAN BEBAS PLAGIASI.....	iii
NOTA DINAS PEMBIMBING	iv
SURAT PERNYATAAN BERJILBAB	v
LEMBAR PERSETUJUAN	vi
LEMBAR PENGESAHAN.....	vii
MOTTO	viii
PERSEMBAHAN	ix
ABSTRAK	x
KATA PENGANTAR	xii
DAFTAR ISI	xiv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xviii
DAFTAR SINGKATAN	xix
BAB I PENDAHULUAN	
A. Latar Belakang Masalah	1
B. Identifikasi Masalah	8
C. Pembatasan Masalah	10
D. Rumusan Masalah	11
E. Tujuan Pengembangan	12
F. Manfaat Pengembangan	13
G. Kajian Penelitian yang Relevan	14
H. Sistematika Pembahasan	18

BAB II LANDASAN TEORI

A. Higher Order Thinking Skills (HOTS)	20
B. Asesmen HOTS dalam Pembelajaran SKI	27
C. Pengembangan Instrumen Asesmen	30
D. Model Rasch	37
E. Kriteria Kelayakan Instrumen Asesmen Berdasarkan Model Rasch....	40
F. Two-Tier Multiple Choice.....	43
G. Pembelajaran SKI di Madrasah Aliyah	46
BAB III METODE PENELITIAN	
A. Jenis Penelitian	52
B. Model Pengembangan	52
C. Subjek Penelitian	54
D. Objek Penelitian	55
E. Desain Uji Coba Produk	57
F. Prosedur Pengembangan	60
G. Teknik dan Instrumen Pengumpulan Data	65
H. Teknik Analisis Data	67
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	
A. Hasil Pengembangan Instrumen TTMC	78
B. Hasil Validasi Ahli dan Uji Coba Produk	82
1. Hasil Validasi Ahli (<i>Expert Review</i>)	83
2. Hasil Uji Coba <i>One to One</i>	86
3. Hasil Uji Coba <i>Small Group</i>	95
C. Analisis Kualitas Instrumen Berdasarkan Model Rasch	100
1. Deskripsi Pelaksanaan <i>Field Test</i>	100
2. Analisis Kualitas Instrumen Berdasarkan Model Rasch	100
D. Analisis Pola Keterpahaman Konsep (K1-K4) Peserta Didik	104

E. Pembahasan	106
1. Proses Pengembangan Instrumen TTMC	106
2. Kualitas Instrumen Berdasarkan Model Rasch	108
3. Keterpahaman Konsep K1 – K4 Peserta Didik	113
F. Revisi Produk Berdasarkan Hasil <i>Field Test</i>	115
G. Analisis Hasil Produk Akhir	117
H. Keterbatasan Penelitian	118
BAB V PENUTUP	
A. Kesimpulan	120
B. Saran	123
DAFTAR PUSTAKA	125
LAMPIRAN	131
DAFTAR RIWAYAT HIDUP.....	232

DAFTAR TABEL

Tabel 1	Integrasi P5-P2RA dalam Tujuan Pembelajaran SKI Kelas XI	
	Materi Dinasti Abbasiyah	56
Tabel 2	Penskoran Keterpahaman TTMC.....	60
Tabel 3	Penskoran Analisis <i>Rasch</i> TTMC	60
Tabel 4	Kisi-kisi Lembar Validitas Ahli	66
Tabel 5	Kriteria Penilaian Tanggapan Validator.....	70
Tabel 6	Kriteria Validitas.....	72
Tabel 7	Kriteria Reliabilitas	73
Tabel 8	Kriteria Daya Beda Berdasarkan Nilai MNSQ	74
Tabel 9	<i>Output</i> Analisis CVI.....	84
Tabel 10	Data Pengolahan Angket Uji Coba <i>One to One</i>	89
Tabel 11	Hasil Keterpahaman Siswa Berdasarkan Uji Small Group.....	95
Tabel 12	Ringkasan Kualitas Butir Soal	103
Tabel 13	Hasil Keterpahaman Siswa Berdasarkan Uji <i>Field Test</i>	104
Tabel 14	Revisi Produk Berdasarkan Uji <i>Field Test</i>	116

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR LAMPIRAN

Lampiran 1	Matriks Analisi CP & ATP	132
Lampiran 2	Kisi-Kisi dan Indikator.....	133
Lampiran 3	<i>Prototype I</i> (Desain Awal)	138
Lampiran 4	Lembar Validasi Ahli Materi	160
Lampiran 5	Lembar Validasi Ahli Evaluasi	163
Lampiran 6	Sampel Angket dan Pertanyaan Terbuka Hasil <i>One to One</i> ..	166
Lampiran 7	Output Analisis Angket Hasil Uji Coba <i>One to One</i>	172
Lampiran 8	Sampel Lembar Observasi dan Wawancara <i>One to One</i>	173
Lampiran 9	Sampel Lembar Jawab Hasil Uji Coba <i>One to One</i>	177
Lampiran 10	Hasil Analisis Jawaban Uji Coba <i>One to One</i>	179
Lampiran 11	Dokumentasi Uji <i>One to One</i>	179
Lampiran 12	Sampel Hasil Angket Uji Coba <i>Small Group</i>	180
Lampiran 13	Output Analisis Angket Hasil Uji Coba <i>Small Group</i>	186
Lampiran 14	Sampel Lembar Jawab Uji Coba <i>Small Group</i>	187
Lampiran 15	Dokumentasi Uji Coba <i>Small Group</i>	191
Lampiran 16	<i>Output</i> Hasil Tes <i>Small Group</i>	192
Lampiran 17	Sampel Lembar Jawab Uji <i>Field Test</i>	194
Lampiran 18	Hasil Olah Data <i>Uji Field Test</i> Menggunakan Jamovi	197
Lampiran 19	<i>Output</i> Hasil Uji <i>Field Test</i> (Keterpahaman Siswa)	199
Lampiran 20	Dokumentasi Uji <i>Field Test</i>	200
Lampiran 21	Analisis Perubahan Tahapan <i>Prototype II, III, dan IV</i>	201
Lampiran 22	<i>Final Product</i>	204
Lampiran 23	Analisis Proporsi Aspek P.A.K pada Instrumen Final	224
Lampiran 24	Panduan Praktis Penyusunan Instrumen Tes TTMC HOTS ...	225

DAFTAR SINGKATAN

TTMC : *Two-Tier Multiple Choice*

HOTS : *Higher Order Thinking Skills*

SKI : Sejarah Kebudayaan Islam

ATP : Alur Tujuan Pembelajaran

TP : Tujuan Pembelajaran

CP : Capaian Pembelajaran



BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Pendidikan abad ke-21 menuntut peserta didik tidak hanya menguasai pengetahuan faktual melalui hafalan, tetapi juga memiliki kemampuan berpikir kritis, analitis, kreatif, serta mampu memecahkan berbagai permasalahan yang dihadapi dalam kehidupan nyata.² Tuntutan tersebut sejalan dengan implementasi Kurikulum Merdeka yang menekankan pengembangan kompetensi peserta didik secara holistik melalui pembelajaran dan asesmen yang berorientasi pada *Higher Order Thinking Skills* (HOTS).³ HOTS merupakan kemampuan berpikir tingkat tinggi yang melibatkan aktivitas menganalisis, mengevaluasi, dan menciptakan solusi berdasarkan informasi yang diperoleh.⁴ Menurut Thomas dan Thorne dalam Nugroho, HOTS tidak hanya menuntut peserta didik mengingat fakta atau menerapkan prosedur yang telah dipelajari, tetapi juga menghubungkan berbagai informasi, menganalisis hubungan antarkonsep, serta menghasilkan gagasan atau solusi baru

² Muh Ridwan Musyaffa and Atno, "Analysis of the History Learning Model Based on Higher Order Thinking Skills (HOTS) in the Merdeka Curriculum," *JPI: Jurnal Pendidikan IPS* Vol. 15, no. 3 (2025): hlm. 731, <https://doi.org/10.37630/jpi.v12i1.617>.

³ Yuniar Handayani, Eni Asia, and Saleh Hidayat, "Peningkatan Kemampuan High Order Thinking Skills (HOTS) Melalui Project-Based Learning (PjBL) dalam Implementasi Kurikulum Merdeka," *PTK: Jurnal Tindakan Kelas* 4, no. 1 (2023): hlm. 50, <https://doi.org/10.53624/ptk.v4i1.236>.

⁴ Hajime Mitani, "Test Score Gaps in Higher Order Thinking Skills: Exploring Instructional Practices to Improve the Skills and Narrow the Gaps," *AERA Open* 7, no. 1 (2021): hlm. 2, <https://doi.org/10.1177/23328584211016470>.

dalam menyelesaikan suatu persoalan.⁵ Secara konseptual, HOTS mengacu pada Taksonomi Bloom Revisi yang menempatkan kemampuan analisis (C4), evaluasi (C5), dan kreasi (C6) sebagai bentuk proses berpikir tingkat tinggi.⁶

Urgensi pengembangan HOTS di Indonesia dapat dilihat dari hasil *Programme for International Student Assessment* (PISA) tahun 2022 yang menunjukkan bahwa kemampuan literasi, numerasi, dan penalaran peserta didik Indonesia masih relatif rendah dibandingkan banyak negara lainnya. Indonesia menempati peringkat ke-68 dari 81 negara peserta. Hasil tersebut menunjukkan bahwa sebagian besar peserta didik masih mengalami kesulitan dalam menyelesaikan soal yang menuntut kemampuan berpikir tingkat tinggi, khususnya kemampuan menganalisis informasi, mengevaluasi berbagai alternatif solusi, dan menggunakan pengetahuan dalam konteks yang baru.⁷ Kondisi ini mengindikasikan bahwa penguatan HOTS masih menjadi tantangan penting dalam sistem pendidikan nasional.⁸

Salah satu faktor yang memengaruhi rendahnya kemampuan HOTS peserta didik adalah sistem asesmen yang belum sepenuhnya

⁵ Arifin Nugroho, *HOTS: Kemampuan Berpikir Tingkat Tinggi* (Jakarta: Gramedia Widiasarana Indonesia, 2018), hlm. 25.

⁶ Lorin W. Anderson and David R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (Longman, 2001).

⁷ Valensy Rachmedita, Risma M. Sinaga, and Pujiati, "Peningkatan Kemampuan Berpikir Kritis Melalui Penggunaan Strategi Active Sharing Knowledge," *Jurnal Studi Sosial* 5, no. 1 (2017): hlm. 3.

⁸ Siti Asfiah, "Implementasi Penilaian Berbasis High Order Thinking Skills pada Mapel PAI dalam Meningkatkan Berpikir Kritis dan Kreatif Siswa di Tingkat SMP," *Quality* 9, no. 1 (2021): hlm. 105.

berorientasi pada pengukuran kemampuan berpikir tingkat tinggi. Annafi menyatakan bahwa kemampuan HOTS tidak dapat diukur secara optimal melalui instrumen asesmen konvensional yang hanya berfokus pada kemampuan mengingat dan memahami informasi faktual.⁹ Namun, dalam praktiknya, asesmen yang digunakan di sekolah masih didominasi oleh tes tertulis yang mengukur kemampuan kognitif tingkat rendah (*Lower Order Thinking Skills/LOTS*). Guru cenderung menggunakan soal yang menekankan aspek mengingat karena lebih mudah disusun dan tersedia dalam berbagai buku teks maupun lembar kerja peserta didik. Selain itu, keterbatasan pemahaman guru mengenai HOTS serta minimnya pelatihan dalam pengembangan instrumen asesmen menjadi faktor yang menyebabkan implementasi asesmen HOTS belum berjalan secara optimal.¹⁰

Padahal, HOTS tidak hanya menjadi tujuan pembelajaran, tetapi juga harus diukur secara terencana melalui instrumen asesmen yang sesuai.¹¹ Asesmen berorientasi HOTS idealnya tidak hanya berfungsi sebagai alat evaluasi hasil belajar, melainkan juga sebagai sarana untuk memperoleh informasi mengenai proses berpikir peserta didik, memberikan umpan balik yang bermakna, mengidentifikasi miskonsepsi,

⁹ Annafi Awantagusnik, "Asesmen Higher Order Thinking Skills (HOTS) Siswa: Pengetahuan dan Praktik Guru Matematika SMP," *SCIENCE: Jurnal Inovasi Pendidikan Matematika Dan IPA* 6, no. 1 (2025): hlm. 531.

¹⁰ Nugroho, *HOTS: Kemampuan Berpikir Tingkat Tinggi*, hlm. 12.

¹¹ Arfan Arfin, Juraid Abdul Latief, and Misnah, "Implementation of HOTS-Based Learning Tools Through a Collaborative Learning Model by Applying the Merdeka Curriculum to Sixth Grade Students at SDN Lahuafu," *Journal of Scientific Research, Education, and Technology (JSRET)* 4, no. 2 (2025): 1239–45, <https://doi.org/10.58526/jsret.v4i2.792>.

dan membantu peserta didik mengembangkan kemampuan berpikir secara berkelanjutan.¹² Oleh karena itu, ketersediaan instrumen asesmen yang mampu mengukur kemampuan berpikir tingkat tinggi secara akurat menjadi kebutuhan yang mendesak dalam praktik pendidikan saat ini.

Kebutuhan tersebut juga terjadi dalam pembelajaran Sejarah Kebudayaan Islam (SKI). Dalam rumpun mata pelajaran Pendidikan Agama Islam (PAI) di Madrasah Aliyah, terdapat lima mata pelajaran utama, yaitu Al-Qur'an Hadis, Akidah Akhlak, Fikih, Sejarah Kebudayaan Islam (SKI), dan Bahasa Arab. Dibandingkan dengan mata pelajaran PAI lainnya, pengembangan instrumen asesmen HOTS berbentuk TTMC pada mata pelajaran SKI jenjang Madrasah Aliyah masih sangat terbatas.¹³ Hal ini menjadikan SKI sebagai fokus yang relevan untuk dikembangkan instrumen asesmennya, khususnya yang mampu mengukur kemampuan berpikir tingkat tinggi peserta didik secara lebih mendalam.

Secara substantif, mata pelajaran SKI memiliki karakteristik yang sangat potensial untuk mengembangkan HOTS karena memuat berbagai peristiwa sejarah, dinamika sosial-politik, perkembangan peradaban Islam, serta nilai-nilai yang dapat dianalisis dan direfleksikan dalam kehidupan masa kini. Peserta didik tidak hanya dituntut mengetahui fakta

¹² Nur Adzidzah and Agus Yudiawan, "HOTS-Based Formative Assessment: The Key to Improving The Quality of Learning," *Journal of Quality Assurance in Islamic Education (JQAIE)* 4, no. 2 (2024): 109–20, <https://doi.org/10.47945/jqaie.v4i2.1670>.

¹³ Muh Ibnu Sholeh et al., "Implementasi Evaluasi Pembelajaran Pendidikan Agama Islam Berbasis Higher Order Thinking Skills (HOTS)," *Al'Ulum Jurnal Pendidikan Islam* 5, no. 1 (2025): 12–26, <https://doi.org/10.54090/alulum.639>.

sejarah, tetapi juga memahami hubungan sebab-akibat suatu peristiwa, mengevaluasi kebijakan tokoh atau pemerintahan tertentu, serta mengambil pelajaran dari peristiwa sejarah untuk menjawab tantangan masa kini. Namun demikian, berbagai penelitian menunjukkan bahwa pembelajaran dan asesmen SKI masih didominasi oleh aktivitas hafalan. Hutami dkk. mengungkapkan bahwa internalisasi nilai-nilai dalam pembelajaran SKI sering kali belum optimal karena pembelajaran lebih berorientasi pada penguasaan fakta sejarah daripada pengembangan kemampuan berpikir kritis.¹⁴ Selain itu, analisis terhadap soal-soal SKI menunjukkan bahwa sebagian besar butir soal masih berada pada level kognitif rendah (C1–C3), sedangkan soal yang mengukur kemampuan analisis dan evaluasi relatif terbatas.¹⁵ Akibatnya, peserta didik cenderung menghafal nama tokoh, tahun, dan peristiwa tanpa memahami makna serta relevansinya dalam kehidupan.

Permasalahan tersebut diperkuat oleh hasil studi pendahuluan yang dilakukan peneliti pada tanggal 18 September 2025 di MAN 2 Sleman melalui wawancara dengan guru mata pelajaran SKI. Hasil wawancara menunjukkan bahwa instrumen asesmen yang digunakan memang telah memuat variasi soal LOTS, MOTS, dan HOTS. Akan tetapi, soal yang secara khusus dirancang untuk mengukur kemampuan

¹⁴ Ingenura Maarti Hutami, Galuh Tisna Widiani, and ul Mardliyah, “Pengembangan Media Komik pada Pembelajaran Sejarah Kebudayaan Islam untuk Meningkatkan Keterampilan Berpikir Kritis Siswa,” *Website: Journal.Unipdu.Ac* 5, no. 1 (2021): hlm. 34.

¹⁵ N Nasifah and I Muliati, “Analisis Soal Higher Order Thinking Skill (HOTS) pada Soal UAS Semester Ganjil Tahun Pelajaran 2020/2021 Mata Pelajaran Sejarah Kebudayaan Islam MA Negeri di Kabupaten Solok Selatan,” *Jurnal Pendidikan Tambusai* 5 (2021): hlm. 5805.

berpikir tingkat tinggi secara mendalam masih sangat terbatas. Instrumen yang digunakan belum mampu menggali alasan atau proses berpikir peserta didik dalam menjawab suatu soal. Selain itu, soal berorientasi HOTS masih belum diberikan secara konsisten dalam proses pembelajaran. Guru juga masih memanfaatkan buku paket Kurikulum 2013 sebagai sumber belajar utama meskipun sekolah telah menerapkan Kurikulum Merdeka. Temuan tersebut menunjukkan bahwa kebutuhan terhadap instrumen asesmen HOTS yang lebih sistematis dan inovatif dalam pembelajaran SKI masih sangat tinggi.

Salah satu alternatif yang dapat digunakan untuk mengatasi permasalahan tersebut adalah pengembangan instrumen asesmen berbentuk *Two-Tier Multiple Choice (TTMC)*. Instrumen ini terdiri atas dua tingkat pertanyaan, yaitu Tier 1 yang berisi pilihan jawaban terhadap suatu permasalahan dan Tier 2 yang berisi pilihan alasan yang mendasari jawaban pada tingkat pertama. Dengan struktur tersebut, TTMC tidak hanya mengukur ketepatan jawaban peserta didik, tetapi juga mampu mengungkap alasan dan proses berpikir yang melatarbelakangi jawaban tersebut. Treagust menjelaskan bahwa instrumen dua tingkat efektif digunakan untuk mengidentifikasi miskonsepsi peserta didik secara lebih akurat dibandingkan tes pilihan ganda konvensional.¹⁶ Dibandingkan instrumen esai yang memerlukan waktu koreksi panjang dan rentan

¹⁶ David F Treagust, "Development and Use of Diagnostic Tests to Evaluate Students' Misconceptions in Science," *International Journal of Science Education* 10, no. 2 (1988): hlm. 163, <https://doi.org/10.1080/0950069880100204>.

subjektivitas,¹⁷ TTMC tetap mempertahankan objektivitas penskoran sekaligus mampu memberikan informasi diagnostik mengenai pemahaman peserta didik.¹⁸ Oleh karena itu, TTMC dipandang relevan untuk digunakan dalam pengukuran HOTS, khususnya pada mata pelajaran SKI yang menuntut kemampuan analisis dan evaluasi terhadap berbagai peristiwa sejarah.

Meskipun demikian, hasil telaah penelitian terdahulu menunjukkan bahwa pengembangan instrumen TTMC berorientasi HOTS pada mata pelajaran SKI jenjang Madrasah Aliyah masih sangat terbatas. Penelitian yang pernah dilakukan dalam pengembangan TTMC dilakukan pada mata pelajaran matematika,¹⁹ sedangkan penelitian pada rumpun Pendidikan Agama Islam masih relatif sedikit dan umumnya dilakukan pada jenjang MTs atau SMP.²⁰ Selain itu, penelitian yang secara khusus mengembangkan instrumen TTMC berorientasi HOTS pada materi Dinasti Abbasiyah kelas XI Madrasah Aliyah masih belum ditemukan. Kesenjangan penelitian lainnya terletak pada penggunaan pendekatan psikometri modern. Sebagian besar penelitian sebelumnya

¹⁷ Norman E. Gronlund and Robert L. Linn, *Measurement and Evaluation in Teaching* (Macmillan Publishing Company, 1990), hlm. 245.

¹⁸ A L Chandrasegaran, D F Treagust, and M Mocerino, "The Development of a Two-tier Multiple-choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Levels of Representation," *Chemistry Education Research and Practice* 8, no. 3 (2007): 293–307, <https://doi.org/10.1039/B7RP90006F>.

¹⁹ Resvia Subay, "Pengembangan Assessment Tes Higher Order Thinking Skills (HOTS) Peserta Didik pada Pembelajaran Matematika Kelas VII Berbasis Model Rasch (Tesis Magister)," *Universitas Negeri Semarang* (Universitas Negeri Semarang, 2020).

²⁰ Muh Ibnu Sholeh et al., "Implementasi Evaluasi Pembelajaran Pendidikan Agama Islam Berbasis Higher Order Thinking Skills (HOTS)," *Al'Ulum Jurnal Pendidikan Islam* 5, no. 1 (2025): 12–26, <https://doi.org/10.54090/alulum.639>.

masih menggunakan analisis tes klasik, sementara pengujian instrumen menggunakan Model Rasch sebagai bagian dari Teori Respons Butir (*Item Response Theory*) masih jarang dilakukan dalam konteks pembelajaran SKI.

Berdasarkan uraian tersebut, diperlukan pengembangan instrumen asesmen HOTS berbentuk *Two-Tier Multiple Choice* pada mata pelajaran Sejarah Kebudayaan Islam materi Dinasti Abbasiyah kelas XI Madrasah Aliyah. Pengembangan instrumen ini diharapkan mampu menghasilkan alat ukur yang valid, reliabel, dan mampu mengidentifikasi proses berpikir serta miskonsepsi peserta didik secara lebih mendalam. Selain itu, penggunaan *Model Rasch* dalam proses analisis instrumen diharapkan dapat memberikan informasi psikometrik yang lebih objektif dan akurat. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan instrumen asesmen HOTS dalam pembelajaran SKI, tetapi juga memperkaya kajian evaluasi pendidikan Islam melalui penerapan pendekatan psikometri modern.

B. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, dapat diidentifikasi beberapa permasalahan utama yang melatarbelakangi penelitian ini sebagai berikut:

1. Dominasi pembelajaran SKI yang berorientasi hafalan menyebabkan internalisasi nilai-nilai kearifan sejarah Islam belum optimal. Peserta

didik kesulitan menemukan korelasi antara materi sejarah dengan kehidupan nyata, diperparah oleh dominasi buku teks yang padat narasi sehingga aktivitas berpikir tingkat tinggi belum berkembang secara maksimal.

2. Meskipun tuntutan pengembangan HOTS telah menjadi orientasi kurikulum yang berlaku, penerapan asesmen yang secara eksplisit mengukur kemampuan berpikir tingkat tinggi dalam pembelajaran SKI masih belum optimal. Penilaian cenderung didominasi oleh tes yang mengukur kemampuan kognitif tingkat rendah, dan kesiapan guru dalam merancang instrumen asesmen HOTS masih menjadi kendala.
3. Instrumen asesmen yang digunakan dalam pembelajaran SKI masih banyak didominasi soal pada level kognitif C1–C3, sementara soal yang menuntut kemampuan analisis (C4) dan evaluasi (C5) relatif terbatas. Temuan pra-observasi di MAN 2 Sleman juga menunjukkan bahwa soal berorientasi HOTS masih jarang diberikan secara konsisten kepada peserta didik.
4. Instrumen asesmen HOTS yang secara khusus dirancang untuk mata pelajaran SKI di jenjang Madrasah Aliyah masih sangat terbatas. Penelitian yang ada umumnya terfokus pada mata pelajaran sains dan matematika, atau dalam rumpun PAI namun terbatas pada jenjang MTs, sehingga terdapat kesenjangan yang perlu diisi.
5. Bentuk asesmen yang umum digunakan hanya menilai jawaban akhir peserta didik tanpa mengetahui alasan atau proses berpikir di baliknya,

sehingga miskonsepsi sulit teridentifikasi. Instrumen konvensional seperti pilihan ganda biasa tidak mampu memotret proses berpikir peserta didik, sementara esai yang ideal untuk HOTS sulit diterapkan secara objektif dalam skala besar.

6. Dari berbagai bentuk instrumen asesmen HOTS yang tersedia, masing-masing memiliki keterbatasan dalam hal objektivitas, efisiensi, dan kemudahan penerapan. Oleh karena itu, diperlukan pengembangan instrumen alternatif yang mampu mengukur HOTS secara komprehensif, objektif, dan terstandarisasi yakni instrumen two-tier multiple choice (TTMC) yang menggabungkan objektivitas pilihan ganda dengan kedalaman penalaran melalui pemilihan alasan pada tingkat kedua.

C. Pembatasan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah diuraikan, maka penelitian ini dibatasi pada pengembangan instrumen asesmen berbasis *Higher Order Thinking Skills* (HOTS) dalam bentuk *two-tier multiple choice* pada mata pelajaran Sejarah Kebudayaan Islam (SKI). Pemilihan mata pelajaran Sejarah Kebudayaan Islam (SKI) sebagai fokus penelitian didasarkan pada beberapa pertimbangan: (1) materi SKI memuat peristiwa-peristiwa sejarah yang kaya akan dimensi kausalitas, kronologi, dan nilai historis sehingga sangat relevan untuk mengukur kemampuan analisis (C4) dan evaluasi (C5); (2) instrumen asesmen HOTS yang secara spesifik menggunakan format *two-tier multiple choice* pada

mata pelajaran SKI di jenjang Madrasah Aliyah masih sangat terbatas; dan (3) hasil studi pendahuluan menunjukkan bahwa instrumen asesmen yang digunakan guru SKI di MAN 2 Sleman belum secara optimal mengukur kemampuan berpikir tingkat tinggi peserta didik. Dengan demikian, instrumen asesmen yang dikembangkan bertujuan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik yang meliputi kemampuan analisis dan evaluasi sesuai dengan taksonomi Bloom Revisi. Penyusunan instrumen asesmen ini mengacu pada Capaian Pembelajaran (CP) mata pelajaran Sejarah Kebudayaan Islam sesuai dengan kurikulum yang berlaku pada satuan pendidikan tersebut. Materi yang digunakan dalam pengembangan instrumen asesmen dibatasi pada materi Dinasti Abbasiyah pada mata pelajaran SKI kelas XI Madrasah Aliyah. Penelitian ini difokuskan pada proses pengembangan serta pengujian kualitas instrumen (*person reliability*, *item fit*, dan *item difficulty*) dan kelayakan instrumen asesmen HOTS berbentuk *two-tier multiple choice*.

D. Rumusan Masalah

Berdasarkan uraian di atas, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana proses pengembangan instrumen tes berbentuk *two-tier multiple choice* berorientasi HOTS pada mata pelajaran Sejarah Kebudayaan Islam materi Dinasti Abbasiyah kelas XI Madrasah Aliyah?
2. Bagaimana tingkat validitas instrumen tes *two-tier multiple choice*

berorientasi HOTS berdasarkan penilaian ahli (*expert judgment*)?

3. Bagaimana kualitas instrumen tes *two-tier multiple choice* berorientasi HOTS ditinjau dari *person reliability*, *item fit*, dan *item difficulty* berdasarkan analisis *Model Rasch* melalui uji coba pada peserta didik kelas XI Madrasah Aliyah?
4. Bagaimana efektivitas instrumen tes *two-tier multiple choice* berorientasi HOTS dalam mengidentifikasi pola keterpahaman konsep (K1-K4) peserta didik kelas XI Madrasah Aliyah pada materi Dinasti Abbasiyah?

E. Tujuan Pengembangan

Berdasarkan rumusan masalah tersebut, maka pengembangan ini bertujuan untuk:

1. Mengembangkan instrumen tes berbentuk *two-tier multiple choice* berorientasi HOTS pada mata pelajaran Sejarah Kebudayaan Islam materi Dinasti Abbasiyah kelas XI Madrasah Aliyah.
2. Mengetahui tingkat validitas instrumen tes *two-tier multiple choice* berorientasi HOTS berdasarkan validasi dari para ahli.
3. Mengetahui kualitas instrumen *two-tier multiple choice* berorientasi HOTS berdasarkan hasil analisis *person reliability*, *item fit*, dan *item difficulty* berdasarkan analisis *Model Rasch* melalui uji coba lapangan.
4. Mendeskripsikan efektivitas instrumen tes *two-tier multiple choice* berorientasi HOTS dalam mengidentifikasi pola keterpahaman

konsep (K1-K4) peserta didik kelas XI Madrasah Aliyah pada materi Dinasti Abbasiyah.

F. Manfaat Pengembangan

Manfaat dari penelitian ini dibagi menjadi:

1. Manfaat Teoretis

Penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan ilmu pendidikan, khususnya dalam bidang evaluasi pembelajaran melalui pengembangan instrumen tes berbentuk *two-tier multiple choice* yang berorientasi HOTS pada mata pelajaran Sejarah Kebudayaan Islam.

2. Manfaat Praktis

a. Bagi Peneliti

Memberikan pengalaman langsung dalam mengaplikasikan model pengembangan instrumen evaluasi yang terstandarisasi, khususnya dalam menyusun soal *two-tier multiple choice* yang valid dan reliabel untuk materi Sejarah Kebudayaan Islam.

b. Bagi Lembaga

Menyediakan instrumen asesmen siap pakai yang telah teruji secara empiris dalam mengevaluasi kecakapan bernalar tingkat tinggi (HOTS) siswa, sekaligus berfungsi sebagai rujukan bagi Madrasah Aliyah dalam memproduksi model evaluasi yang sejenis.

c. Bagi Peserta Didik

Memfasilitasi peserta didik untuk mengidentifikasi tingkat pemahaman konsep dan miskonsepsi melalui struktur soal dua tingkat (*two-tier*), sekaligus membiasakan diri peserta didik menghadapi instrumen evaluasi berorientasi HOTS.

d. Bagi Peneliti Berikutnya

Memberikan data dasar dan kerangka kerja mengenai efektivitas penggunaan soal dua tingkat (*two-tier*) pada mata pelajaran rumpun PAI, yang dapat dikembangkan lebih lanjut pada cakupan materi atau variabel penelitian yang lebih luas.

G. Kajian Penelitian yang Relevan

Pengembangan instrumen penilaian berorientasi *Higher Order Thinking Skills* (HOTS) telah menjadi perhatian serius dalam berbagai penelitian pendidikan Islam. Beberapa penelitian telah mengembangkan instrumen HOTS berbentuk *two-tier multiple choice* (TTMC) dalam konteks madrasah, di antaranya Azty,²¹ Alhadad,²² Nurmayanti dan

²¹ Alnida Azty, "Analisis Kualitas Butir Soal Pendidikan Agama Islam (Tesis Magister)" (Universitas Islam Negeri Sumatera Utara, 2020).

²² Ahmad Alhadad, "Pengembangan Instrumen Tes Fikih MTs Bab Zakat Berorientasi Higher Order Thinking Skills (HOTS) di MTs N 4 Sleman (Tesis Magister)" (Universitas Islam Negeri Sunan Kalijaga, 2022).

Basid,²³ Humaidi dkk.,²⁴ Sholeh dkk.,²⁵ Pramudita dkk.,²⁶ serta Andriyatno dkk.²⁷ Penelitian lain mengembangkan instrumen HOTS dengan format berbeda, Jayanti menggunakan pilihan ganda konvensional dengan validasi Aiken's V dan CFA pada jenjang SD,²⁸ sementara Abidin dkk. mengembangkan instrumen evaluasi autentik multiformat (kinerja, portofolio, dan tes objektif) pada jenjang SMP.²⁹ Secara umum, penelitian-penelitian tersebut menunjukkan bahwa instrumen TTMC terbukti efektif untuk mengukur kemampuan analisis dan evaluasi peserta didik sekaligus mengungkap pola pemahaman dan miskonsepsi yang tidak terdeteksi oleh instrumen pilihan ganda konvensional. Instrumen yang dikembangkan dalam penelitian-penelitian tersebut dinyatakan valid, reliabel, dan layak digunakan dalam konteks pembelajaran PAI di madrasah.

Namun demikian, penelitian-penelitian tersebut memiliki

²³ Iin Nurmayanti and Abdul Basid, "Efektivitas Instrumen Uji Soal HOTS dan Budaya Pengambilan Nilai Ulangan Harian pada Madrasah," *Penamas* 34, no. 2 (2021): 335–52, <https://doi.org/10.31330/penamas.v34i2.516>.

²⁴ M. Nurul Humaidi et al., "Development of a HOTS-Leveled Two-Tier Multiple Choice (TTMC) Test to Measure Student Misconceptions in Islamic Studies," *Assyfa Journal of Islamic Studies* 1, no. 1 (2023): 31–40, <https://doi.org/10.61650/ajis.v1i1.148>.

²⁵ Muh Ibnu Sholeh et al., "Implementasi Evaluasi Pembelajaran Pendidikan Agama Islam Berbasis Higher Order Thinking Skills (HOTS)," *Al'Ulum Jurnal Pendidikan Islam* 5, no. 1 (2025): 12–26, <https://doi.org/10.54090/alulum.639>.

²⁶ Arya Pramudita et al., "Implementasi Instrumen Tes Objektif Berbasis HOTS dalam Evaluasi Pembelajaran PAI di SMA Negeri 1 Jetis Bantul," *Reflection : Islamic Education Journal* 2, no. 4 (2025): 01–18, <https://doi.org/10.61132/reflection.v2i4.1326>.

²⁷ Indri Andriyatno, Zulfiani Zulfiani, and Yuke Mardiaty, "Higher Order Thinking Skills: Student Profile Using Two-Tier Multiple Choice Instrument," *International Journal of STEM Education for Sustainability* 3, no. 1 (2023): 111–24, <https://doi.org/10.53889/ijses.v3i1.79>.

²⁸ Dwi Jayanti, "Pengembangan Instrumen Tes Higher Order Thinking Skills (HOTS) pada Mata Pelajaran Pendidikan Agama Islam Kelas 5 di Sekolah Dasar Negeri Kebon Jeruk Jakarta Barat (Tesis Magister)" (UIN Syarif Hidayatullah, 2023).

²⁹ Muhammad Arif Zaenal Abidin and Ahmad Fathi Naufal Hasby, "Pengembangan Instrumen Evaluasi Otentik Berbasis HOTS pada Pembelajaran Pendidikan Agama Islam di SMP Darul Huda Banyuwangi," *MADRASA: Jurnal Pendidikan dan Teknologi* 2, no. 1 (2026): 2–11.

keterbatasan yang perlu dicermati. Pertama, hampir seluruhnya dilakukan pada jenjang MTs atau SMP, dengan cakupan materi PAI secara umum atau fikih, belum ada yang secara spesifik menyorot jenjang Madrasah Aliyah pada mata pelajaran Sejarah Kebudayaan Islam (SKI). Kedua, sebagian besar penelitian hanya melakukan validasi ahli dan uji coba terbatas tanpa analisis psikometrik yang komprehensif, sehingga kualitas butir soal secara empiris belum dapat diverifikasi secara mendalam. Ketiga, tidak ada yang mengintegrasikan analisis Model Rasch dalam proses pengembangan instrumen TTMC HOTS pada konteks PAI maupun SKI.

Dalam konteks SKI, beberapa penelitian relevan telah dilakukan, namun dengan pendekatan yang berbeda. Hoerul mengembangkan bahan ajar HOTS untuk SKI kelas VIII MTs,³⁰ sementara Harahap mengembangkan instrumen HOTS berbentuk esai untuk SKI kelas X MA dengan hasil validitas dan praktikalitas yang baik.³¹ Khairunamira dan Rohimah menemukan bahwa sebagian besar soal SKI di lapangan masih tergolong LOTS dan guru mengalami kesulitan merancang soal HOTS.³² Temuan-temuan ini menegaskan bahwa meskipun urgensi

³⁰ Hoerul Ansori, "Pengembangan Bahan Ajar Berbasis Higher Order Thinking Skills (HOTS) pada Mata Pelajaran Sejarah Kebudayaan Islam Kelas VIII MTsN 3 Sleman Semester Ganjil (Tesis Magister)" (UIN Sunan Kalijaga Yogyakarta, 2019).

³¹ Norlan Cahaya Harahap, "Pengembangan Instrumen Asesmen Berbasis HOTS Kompetensi Pengetahuan pada Mata Pelajaran Sejarah Kebudayaan Islam di Kelas X Madrasah Aliyah Negeri 3 Tapanuli Tengah (Tesis Magister)" (UIN Syekh Ali Hasan Ahmad Addary Padangsidempuan, 2024).

³² Fathya Khairunamira and Siti Rohimah, "Evaluation of HOTS-Based Questions in Islamic Cultural History at Insan Cendekia Boarding School Sukoharjo," *Asatiza: Jurnal Pendidikan* 6, no. 1 (2025): 88–101.

pengembangan instrumen HOTS dalam SKI telah disadari, upaya pengembangan instrumen yang terstruktur, terstandar, dan bersifat diagnostik khususnya dalam format TTMC masih sangat terbatas.

Di sisi lain, beberapa penelitian telah memanfaatkan analisis Model Rasch dan *Item Response Theory* (IRT) dalam evaluasi kualitas instrumen. Prihatin menganalisis kualitas soal PAT SKI kelas X MAN 2 Bantul menggunakan IRT dan menemukan dominasi soal LOTS (C1–C2).³³ Pitriani dkk. menganalisis 160 butir soal UAS PAI di MTs menggunakan Model Rasch dan menemukan 75% soal masih berada pada level LOTS dengan 37,5% butir tidak fit.³⁴ Subay mengembangkan instrumen HOTS matematika kelas VII menggunakan Model Rasch dengan format esai.³⁵ Penelitian-penelitian tersebut membuktikan bahwa Model Rasch mampu memberikan gambaran kualitas instrumen yang lebih komprehensif dibandingkan pendekatan teori tes klasik. Akan tetapi, tidak satu pun dari penelitian tersebut yang mengintegrasikan pengembangan instrumen TTMC HOTS dengan analisis Model Rasch pada mata pelajaran SKI di jenjang Madrasah Aliyah.

Berdasarkan pemetaan tersebut, terdapat tiga *research gap* yang menjadi fokus penelitian ini. Pertama, belum ditemukan penelitian yang

³³ Rani Putri Prihatin, “Analisis Soal SKI Kelas X MAN 2 Bantul Yogyakarta pada Penilaian Akhir Tahun 2020/2021 (Tesis Magister)” (UIN Sunan Kalijaga, 2022).

³⁴ Pipit Pitriani et al., “Identification of HOTS-Based Islamic Education Tests in Junior High Schools : An Approach Bloom’s Taxonomy and Rasch Model,” *Al-Ishlah: Jurnal Pendidikan Islam* 22, no. 1 (2024): 30–44, <https://doi.org/10.35905/alishlah.v22i1.9649>.

³⁵ Resvia Subay, “Pengembangan Assessment Tes Higher Order Thinking Skills (HOTS) Peserta Didik pada Pembelajaran Matematika Kelas VII Berbasis Model Rasch (Tesis Magister),” *Universitas Negeri Semarang* (Universitas Negeri Semarang, 2020).

mengembangkan instrumen HOTS berbentuk *two-tier multiple choice* secara spesifik pada mata pelajaran SKI di tingkat Madrasah Aliyah. Kedua, belum ada penelitian yang secara khusus mengembangkan instrumen HOTS pada materi Dinasti Abbasiyah, materi yang kaya dimensi kausalitas, kronologis, dan evaluatif sehingga relevan untuk mengukur kemampuan berpikir tingkat tinggi. Ketiga, integrasi pengembangan instrumen TTMC dengan analisis Model Rasch dalam konteks pembelajaran SKI masih sangat terbatas. Kebaruan (*novelty*) penelitian ini terletak pada upaya mengisi ketiga celah tersebut melalui pengembangan instrumen asesmen HOTS berbentuk *two-tier multiple choice* pada mata pelajaran SKI kelas XI Madrasah Aliyah materi Dinasti Abbasiyah yang dianalisis menggunakan Model Rasch, sehingga instrumen yang dihasilkan tidak hanya terukur secara konten, tetapi juga terverifikasi secara psikometrik.

H. Sistematika Pembahasan

Sistematika ini disusun guna menyajikan alur pembahasan penelitian secara terstruktur. Adapun sistematika dalam penelitian ini sebagai berikut:

Bab I: Pendahuluan, yang berisi; Latar Belakang Masalah, Identifikasi Masalah, Pembatasan Masalah, Rumusan Masalah, Tujuan Pengembangan, Manfaat Pengembangan, Kajian Penelitian yang Relevan, dan Sistematika Pembahasan.

- Bab II: Landasan Teori
- Bab III: Metode Penelitian, yang berisi; Jenis Penelitian, Model Pengembangan, Subjek Penelitian, Objek Penelitian, Desain Uji Coba Produk, Prosedur Pengembangan, Teknik dan Instrumen Pengumpulan Data, dan Teknik Analisis Data.
- Bab IV: Hasil Penelitian dan Pembahasan, yang menyajikan; Hasil Pengembangan Produk Awal, Hasil Uji Coba Produk, Revisi Produk, Analisis Hasil Produk Akhir dan Keterbatasan Penelitian.
- Bab V: Penutup, yang berisi; Simpulan dan Saran.



BAB V PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian dan pengembangan yang telah dilaksanakan, dapat ditarik kesimpulan sebagai berikut:

1. Instrumen asesmen *Two-Tier Multiple Choice* (TTMC) berorientasi HOTS pada mata pelajaran SKI materi Dinasti Abbasiyah dikembangkan melalui model *formative evaluation* Tessmer yang mencakup empat tahapan: (1) tahap *preliminary* berupa studi pendahuluan di MAN 2 Sleman; (2) tahap *self-evaluation* yang meliputi analisis kurikulum Merdeka, analisis peserta didik, analisis materi Dinasti Abbasiyah, dan penyusunan desain instrumen (*Prototype I*); (3) tahap *prototype I* melalui validasi ahli (*expert review*), uji *one-to-one* pada 6 peserta didik, dan uji *small group* pada 15 peserta didik; serta (4) tahap *field test* pada 30 peserta didik kelas XI MAN 2 Sleman. Produk akhir berupa 30 butir soal TTMC yang terdiri dari *tier 1* (pilihan jawaban) dan *tier 2* (alasan) dengan sebaran level kognitif C4 (analisis) dan C5 (evaluasi).
2. Berdasarkan penilaian dua validator ahli (ahli materi dan ahli evaluasi), instrumen TTMC HOTS yang dikembangkan memiliki validitas isi yang baik. Hasil perhitungan *Content Validity Index* (CVI) menunjukkan bahwa 10 dari 13 indikator memperoleh nilai I-CVI sebesar 1,00 (kesepakatan sempurna), sedangkan 3 indikator

memperoleh nilai I-CVI sebesar 0,50 dan telah diperbaiki sesuai saran validator. Nilai S-CVI/*Average* keseluruhan instrumen mencapai 0,88, yang mengindikasikan bahwa instrumen memiliki validitas isi yang baik dan layak digunakan, mengacu pada kriteria Polit & Beck ($\geq 0,90$ sangat baik; $\geq 0,80$ baik).

3. Hasil analisis *field test* menggunakan *Model Rasch* menunjukkan bahwa:
 - (a) *Person Reliability*. Nilai *Person Reliability* sebesar 0,629 berada pada kategori cukup. Nilai ini dipengaruhi oleh keterbatasan jumlah sampel (30 responden) dan kemungkinan adanya *test anxiety* saat menghadapi format TTMC yang relatif baru bagi peserta didik.
 - (b) *Item Fit* (Validitas Butir dan Daya Beda). Dari 30 butir soal, sebanyak 24 (80%) item menunjukkan nilai Infit dan Outfit MNSQ dalam rentang 0,5 – 1,5, sehingga dinyatakan valid dan memiliki daya beda yang baik (*productive for measurement*). Satu item mengalami misfit yaitu item P (Outfit = 2,531) yang menunjukkan pola jawaban tidak terprediksi akibat distraktor yang kurang efektif. Sebanyak 5 item mengalami overfit (Outfit < 0,5), yaitu item U, V, W, X, dan AF, yang menandakan soal terlalu mudah ditebak sehingga kurang memberikan informasi pembeda yang bermakna. Khusus item AF dijawab benar oleh seluruh responden (*Proportion* = 1,000) sehingga tidak memiliki fungsi diskriminasi sama sekali.
 - (c) *Item Difficulty* (Tingkat Kesukaran). Berdasarkan *Wright Map*, sebagian besar butir soal terpusat pada logit negatif, mengindikasikan

bahwa instrumen masih tergolong mudah bagi peserta didik sehingga belum optimal dalam mengukur kemampuan HOTS secara tajam. Butir dengan tingkat kesukaran tertinggi adalah item P (logit 2,034), G (1,758), dan F (1,311), yang lebih sesuai dengan karakteristik soal HOTS.

Setelah dilakukan revisi produk akhir melalui perbaikan item misfit (P), peningkatan kompleksitas item AF, dan kalibrasi ulang distraktor pada item overfit (U, V, W, X,), instrumen TTMC dinilai berpotensi layak sebagai alat evaluasi diagnostik untuk mengukur kemampuan berpikir tingkat tinggi peserta didik MA kelas XI pada materi Dinasti Abbasiyah.

4. Instrumen asesmen TTMC yang dikembangkan terbukti efektif sebagai instrumen diagnostik karena mampu memetakan profil kognitif siswa secara mendalam. Efektivitas ini ditunjukkan oleh data hasil identifikasi pola keterpahaman, yaitu: kategori K1 (paham konsep) sebesar 69,4% mengindikasikan bahwa sebagian besar peserta didik mampu menjawab dengan benar disertai alasan yang tepat; K2 (paham parsial) sebesar 8,1%; K3 (miskonsepsi) sebesar 9,4%; dan K4 (tidak paham) sebesar 13,0%. Teridentifikasinya kategori miskonsepsi dan tidak paham secara spesifik membuktikan bahwa instrumen TTMC mampu mengungkap informasi diagnostik yang tidak dapat diperoleh melalui instrumen pilihan ganda konvensional, sehingga dapat digunakan guru sebagai dasar tindak lanjut pembelajaran yang lebih terarah.

B. Saran

Berdasarkan hasil penelitian dan keterbatasan yang telah diuraikan, peneliti mengajukan beberapa saran sebagai berikut:

1. Bagi Guru SKI

Instrumen TTMC ini dapat digunakan sebagai alternatif soal ulangan harian atau penilaian sumatif yang mampu menggali proses berpikir peserta didik sekaligus mengidentifikasi miskonsepsi, bukan sekadar menilai jawaban akhir. Guru disarankan memperhatikan hasil analisis keterpahaman (K1–K4) sebagai dasar tindak lanjut pembelajaran.

2. Bagi Sekolah (MAN 2 Sleman)

Disarankan untuk menjadikan instrumen ini sebagai referensi dalam pengembangan bank soal HOTS pada mata pelajaran rumpun PAI lainnya, serta mendorong penggunaan analisis Model Rasch sebagai evaluasi kualitas instrumen di lingkungan madrasah.

3. Bagi Peneliti Berikutnya

Perlu dilakukan uji coba dengan jumlah sampel yang lebih besar (minimal 100–200 responden) agar analisis *Model Rasch* menghasilkan estimasi parameter yang lebih stabil dan akurat. Selain itu, mengingat penelitian ini secara spesifik berfokus pada mata pelajaran SKI, peneliti berikutnya disarankan untuk mengembangkan instrumen serupa pada mata pelajaran rumpun PAI lainnya di Madrasah Aliyah, seperti Al-Qur'an Hadis, Akidah Akhlak, dan Fikih, sehingga ketersediaan instrumen asesmen HOTS berbentuk TTMC dapat merata di seluruh

mata pelajaran agama Islam. Peneliti berikutnya juga disarankan untuk memperhatikan distribusi tingkat kesukaran soal sejak tahap desain agar *Wright Map* menghasilkan sebaran item yang lebih merata.

4. Bagi Pengembang Instrumen

Perlu dilakukan kalibrasi distraktor yang lebih cermat sejak tahap desain, khususnya pada tier 2 (alasan), agar tingkat kesukaran soal terdistribusi merata dari mudah hingga sulit. Pengembang juga disarankan untuk memastikan tidak ada butir soal yang dijawab benar oleh seluruh responden ($Proportion = 1,000$) karena hal tersebut menandakan soal tidak memiliki fungsi diskriminasi sama sekali dalam konteks pengukuran HOTS.

5. Bagi Guru yang Ingin Mengembangkan Instrumen Serupa

Sebagai tindak lanjut praktis dari penelitian ini, peneliti telah menyusun panduan praktis penyusunan instrumen TTMC berorientasi HOTS yang dapat digunakan langsung oleh guru SKI tanpa harus membaca keseluruhan tesis ini. Panduan tersebut mencakup langkah-langkah penyusunan kisi-kisi, penulisan soal tier 1 dan tier 2, lembar telaah mandiri, serta interpretasi hasil K1–K4 untuk keperluan diagnostik (lihat Lampiran 24).

DAFTAR PUSTAKA

- Abdurrahman, Dudung. *Metodologi Penelitian Sejarah*. Yogyakarta: Ar-Ruzz Media, 2007.
- Abidin, Muhammad Arif Zaenal, and Ahmad Fathi Naufal Hasby. "Pengembangan Instrumen Evaluasi Otentik Berbasis HOTS pada Pembelajaran Pendidikan Agama Islam di SMP Darul Huda Banyuwangi." *MADRASA: Jurnal Pendidikan dan Teknologi* 2, no. 1 (2026): 2–11.
- Adnannudin, Adnannudin, Suherli Kusmana, and Dede Endang Mascita. "Pengembangan Penilaian Kognitif Berorientasi HOTS dan Pemanfaatannya dalam Pembelajaran Bahasa Indonesia di SMK." *Jurnal Tuturan* 9, no. 1 (2020): 1. <https://doi.org/10.33603/jt.v9i1.3635>.
- Adzidzah, Nur, and Agus Yudiawan. "HOTS-Based Formative Assessment: The Key to Improving The Quality of Learning." *Journal of Quality Assurance in Islamic Education (JQAIE)* 4, no. 2 (2024): 109–20. <https://doi.org/10.47945/jqaie.v4i2.1670>.
- Afifah Aulia Zayrin, Hayatun Nupus, Khalista Khansa Maizia, Siska Marsela, Rully Hidayatullah, and Harmonedi Harmonedi. "Analisis Instrumen Penelitian Pendidikan (Uji Validitas dan Relibilitas Instrumen Penelitian)." *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora* 3, no. 2 (2025): 780–89. <https://doi.org/10.61104/jq.v3i2.1070>.
- Akbar, Rahmadani, Muhammad Fadjri, and Suwadi. "Analisis Validitas Instrumen Asesmen pada Materi Tokoh-Tokoh Gerakan Pembaruan Islam dalam Buku SKI Kelas 12 MA Muallimin Muhammadiyah Yogyakarta." *Jurnal Ilmiah Pendidikan Dasar* 10, no. 2 (2025): 419–39.
- Alhadad, Ahmad. "Pengembangan Instrumen Tes Fikih MTs Bab Zakat Berorientasi Higher Order Thinking Skills (HOTS) di MTs N 4 Sleman (Tesis Magister)." Universitas Islam Negeri Sunan Kalijaga, 2022.
- Alnida Azty. "Analisis Kualitas Butir Soal Pendidikan Agama Islam (Tesis Magister)." Universitas Islam Negeri Sumatera Utara, 2020.
- Anderson, Lorin W., and David R. Krathwohl. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman, 2001.
- Andriyatno, Indri, Zulfiani Zulfiani, and Yuke Mardiaty. "Higher Order Thinking Skills: Student Profile Using Two-Tier Multiple Choice Instrument." *International Journal of STEM Education for Sustainability* 3, no. 1 (2023): 111–24. <https://doi.org/10.53889/ijses.v3i1.79>.
- Ansori, Hoerul. "Pengembangan Bahan Ajar Berbasis Higher Order Thinking Skills (HOTS) pada Mata Pelajaran Sejarah Kebudayaan Islam Kelas VIII MTsN 3 Sleman Semester Ganjil (Tesis Magister)." UIN Sunan Kalijaga Yogyakarta,

2019.

- Arfin, Arfan, Juraid Abdul Latief, and Misnah. "Implementation of HOTS-Based Learning Tools Through a Collaborative Learning Model by Applying the Merdeka Curriculum to Sixth Grade Students at SDN Lahuaflu." *Journal of Scientific Research, Education, and Technology (JSRET)* 4, no. 2 (2025): 1239–45. <https://doi.org/10.58526/jsret.v4i2.792>.
- Ariyana, Yoki. *Buku Pegangan Pembelajaran Berorientasi pada Keterampilan Berpikir Tingkat Tinggi*. Jakarta: Direktorat Jendral Guru dan Tenaga Kependidikan Kementerian dan Kebudayaan, 2018.
- Arya Pramudita, Wahyu Kholis Prihantoro, Naqiyatul Fikriyah, Luthfi Falasifah, and Safa'atun Nursolihah. "Implementasi Instrumen Tes Objektif Berbasis HOTS dalam Evaluasi Pembelajaran PAI di SMA Negeri 1 Jetis Bantul." *Reflection: Islamic Education Journal* 2, no. 4 (2025): 01–18. <https://doi.org/10.61132/reflection.v2i4.1326>.
- Asfiyah, Siti. "Implementasi Penilaian Berbasis High Order Thinking Skills pada Mapel PAI dalam Meningkatkan Berpikir Kritis dan Kreatif Siswa di Tingkat SMP." *Quality* 9, no. 1 (2021): 103–20.
- Awantagusnik, Annafi. "Asesmen Higher Order Thinking Skills (HOTS) Siswa: Pengetahuan dan Praktik Guru Matematika SMP." *SCIENCE: Jurnal Inovasi Pendidikan Matematika dan IPA* 6, no. 1 (2025): 530–40.
- Bintang, Rahmat S., and Suprananto. "The Impact of Sample Size, Test Length, and Person-Item Targeting on The Separation Reliability in Rasch Model: A Simulation Study." *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia* 13, no. 1 (2024): 1–13. <https://doi.org/10.15408/jp3i.v13i1.27975>.
- Bond, Trevor G., Zheng Yan, and Moritz Heene. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 4th ed. New York: Routledge, 2021.
- Brookhart, Susan M. *How to Assess Higher-Order Thinking Skills in Your Classroom*. ASCD, 2010.
- Brown, James Dean, and Thom Hudson. *The Criterion-Referenced Test Guide: Developing and Interpreting Validity and Reliability Measures*. Greenwood Publishing Group, 2006.
- Chandrasegaran, A L, D F Treagust, and M Mocerino. "The Development of a Two-tier Multiple-choice Diagnostic Instrument for Evaluating Secondary School Students' Ability to Describe and Explain Chemical Reactions Using Multiple Levels of Representation." *Chemistry Education Research and Practice* 8, no. 3 (2007): 293–307. <https://doi.org/10.1039/B7RP90006F>.
- Darmalinda, and Fadriati. "Learning History of Islamic Culture (Analysis of Concepts, Objectives, Materials, Strategies, and Learning Evaluation)." *Al-Hasanah: Jurnal Pendidikan Agama Islam* 9, no. 1 (2024): 92–107.

- Ekstrand, Joakim, Albert Westergren, Kristofer Årestedt, Amanda Hellström, and Peter Hagell. "Transformation of Rasch Model Logits for Enhanced Interpretability." *BMC Medical Research Methodology* 22, no. 1 (2022): 1–10. <https://doi.org/10.1186/s12874-022-01816-1>.
- Erlinawati, Erlinawati, and Muslimah Muslimah. "Test Validity and Reliability in Learning Evaluation." *Bulletin of Community Engagement* 1, no. 1 (2021): 26. <https://doi.org/10.51278/bce.v1i1.96>.
- Gronlund, Norman E., and Robert L. Linn. *Measurement and Evaluation in Teaching*. Macmillan Publishing Company, 1990.
- Handayani, Yuniar, Eni Asia, and Saleh Hidayat. "Peningkatan Kemampuan High Order Thinking Skills (HOTS) Melalui Project-Based Learning (PjBL) dalam Implementasi Kurikulum Merdeka." *PTK: Jurnal Tindakan Kelas* 4, no. 1 (2023): 48–60. <https://doi.org/10.53624/ptk.v4i1.236>.
- Harahap, Norlan Cahaya. "Pengembangan Instrumen Asesmen Berbasis HOTS Kompetensi Pengetahuan pada Mata Pelajaran Sejarah Kebudayaan Islam di Kelas X Madrasah Aliyah Negeri 3 Tapanuli Tengah (Tesis Magister)." UIN Syekh Ali Hasan Ahmad Addary Padangsidempuan, 2024.
- Humaidi, M. Nurul, Fadli Agus Triansyah, Rahmad Sugianto, and Alfi Rachma Nisfi Laila. "Development of a HOTS-Leveled Two-Tier Multiple Choice (TTMC) Test to Measure Student Misconceptions in Islamic Studies." *Assyfa Journal of Islamic Studies* 1, no. 1 (2023): 31–40. <https://doi.org/10.61650/ajis.v1i1.148>.
- Indrawati, Nenny, and Eni Elawati. "Pengembangan Instrumen Tes Higher Order Thinking Skill (HOTS) pada Materi Statistika SMP." *Jurnal Pendidikan dan Konseling* 4, no. 6 (2022): 4458–69.
- Iqlima Zahari. "Implementasi Metode Diskusi dalam Meningkatkan Motivasi Belajar Siswa pada Mata Pelajaran Sejarah Kebudayaan Islam (SKI) Kelas X MAN 5 Kediri." *ENGANG: Jurnal Pendidikan, Bahasa, Sastra, Seni, dan Budaya* 2, no. 2 (2022): 408–15. <https://doi.org/10.37304/enggang.v2i2.12758>.
- Jayanti, Dwi. "Pengembangan Instrumen Tes Higher Order Thinking Skill (HOTS) pada Mata Pelajaran Pendidikan Agama Islam Kelas 5 di Sekolah Dasar Negeri Kebon Jeruk Jakarta Barat (Tesis Magister)." UIN Syarif Hidayatullah, 2023.
- Khairunamira, Fathya, and Siti Rohimah. "Evaluation of HOTS-Based Questions in Islamic Cultural History at Insan Cendekia Boarding School Sukoharjo." *Asatiza: Jurnal Pendidikan* 6, no. 1 (2025): 88–101.
- Kristanto, Priantoro Dwi, and Paula Glady Frandani Setiawan. *Pengembangan Soal HOTS (Higher Order Thinking Skills) Terkait dengan Konteks Pedesaan. PRISMA: Prosiding Seminar Nasional Matematika*. Vol. 3, 2020.

<https://journal.unnes.ac.id/sju/index.php/prisma/article/view/37616>.

Kurniawan, Heru. *Pengantar Praktis Penyusunan Instrumen Penelitian*. Yogyakarta: Deepublish, 2021.

Lynn, Mary R. "Determination and Quantification of Content Validity." *Nursing Research* 35, no. 6 (1986): 382–85.

Maarti Hutami, Ingenura, Galuh Tisna Widiana, and ul Mardiyah. "Pengembangan Media Komik pada Pembelajaran Sejarah Kebudayaan Islam untuk Meningkatkan Keterampilan Berpikir Kritis Siswa." *Website: Journal.Unipdu.Ac* 5, no. 1 (2021): 33–50.

Masitoh, Lisda Fitriana, and Weni Gurita Aedi. "Pengembangan Instrumen Asesmen Higher Order Thinking Skills (HOTS) Matematika di SMP Kelas VII." *Jurnal Cendekia : Jurnal Pendidikan Matematika* 4, no. 2 (2020): 886–97. <https://doi.org/10.31004/cendekia.v4i2.328>.

McMillan, James H. *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction*. 5th ed. Pearson Education, 2011.

Mitani, Hajime. "Test Score Gaps in Higher Order Thinking Skills: Exploring Instructional Practices to Improve the Skills and Narrow the Gaps." *AERA Open* 7, no. 1 (2021): 1–23. <https://doi.org/10.1177/23328584211016470>.

Musyaffa, Muh Ridwan, and Atno. "Analysis of the History Learning Model Based on Higher Order Thinking Skills (HOTS) in the Merdeka Curriculum." *JPI: Jurnal Pendidikan IPS* Vol. 15, no. 3 (2025): 730–39. <https://doi.org/10.37630/jpi.v12i1.617>.

Nasifah, N, and I Muliati. "Analisis Soal Higher Order Thinking Skill (HOTS) pada Soal UAS Semester Ganjil Tahun Pelajaran 2020/2021 Mata Pelajaran Sejarah Kebudayaan Islam MA Negeri di Kabupaten Solok Selatan." *Jurnal Pendidikan Tambusai* 5 (2021): 5804–14.

Nirmala, Z, Remiswal, and Khadijah. "Analisis Soal Asesmen Sumatif Pembelajaran Fiqih Ditinjau Berdasarkan Tipe HOTS Menggunakan Taksonomi Bloom." *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia* 14, no. 1 (2024): 11–20.

Nugroho, Arifin. *HOTS: Kemampuan Berpikir Tingkat Tinggi*. Jakarta: Gramedia Widiasarana Indonesia, 2018.

Nunnally, Jum C., and Ira H. Bernstein. *Psychometric Theory*. McGraw-Hill, 1994.

Nurmayanti, Iin, and Abdul Basid. "Efektivitas Instrumen Uji Soal HOTS dan Budaya Pengambilan Nilai Ulangan Harian pada Madrasah." *Penamas* 34, no. 2 (2021): 335–52. <https://doi.org/10.31330/penamas.v34i2.516>.

Pitriani, Pipit, Ilzamudin Ma'mur, Enung Nugraha, Washudin, and Muhammad Habibi. "Identification of HOTS-Based Islamic Education Tests in Junior High Schools : An Approach Bloom's Taxonomy and Rasch Model." *Al-*

- Ishlah: Jurnal Pendidikan Islam* 22, no. 1 (2024): 30–44. <https://doi.org/10.35905/alishlah.v22i1.9649>.
- Plutarch. “On Listening to Lectures (Trans. Robin Waterfield).” In *Essays by Plutarch*, 3–18. London: Penguin Classics, 1992.
- Polit, Denise F., and Cheryl Tatano Beck. *Essentials of Nursing Research: Methods, Appraisal, and Utilization*. Philadelphia, PA: Lippincott Williams & Wilkins, 2006.
- Prihatin, Rani Putri. “Analisis Soal SKI Kelas X MAN 2 Bantul Yogyakarta pada Penilaian Akhir Tahun 2020/2021 (Tesis Magister).” UIN Sunan Kalijaga, 2022.
- Rachmedita, Valensy, Risma M. Sinaga, and Pujiati. “Peningkatan Kemampuan Berpikir Kritis Melalui Penggunaan Strategi Active Sharing Knowledge.” *Jurnal Studi Sosial* 5, no. 1 (2017).
- Rahman, Abdul, Ni Made Wirastika Sari, Mochamad Sugiarto, Zainal Abidin, and Anton Priyo Nugroho. *Metode Penelitian Ilmu Sosial*. Bandung: Widina Bhakti Persada, 2022.
- Rukayah, Jenny Indrastoeti Siti Poerwanti, and Hafizah Lukitasari. *Penyusunan Two Tier Multiple Choice Test untuk Mengukur Higher Order Thinking Skill (HOTS) di Sekolah Dasar*. Surakarta: UNS Press, 2018.
- Sari Putri, Meilinda, and Hery Noer Aly. “Taksonomi Tujuan Pendidikan dalam Kurikulum Merdeka pada Pembelajaran Pendidikan Agama Islam di Sekolah Dasar Islam Terpadu Kota Bengkulu.” *Science and Education Journal (SICEDU)* 2, no. 2 (2023): 356–63. <https://doi.org/10.31004/sicedu.v2i2.127>.
- Setiawati, Wiwik, Oktavia Asmira, Yoki Ariyana, Reisky Bestary, and Ari Pudjiastuti. *Buku Penilaian Berorientasi Higher Order Thinking Skills. Direktorat Jenderal Guru dan Tenaga Kependidikan Kementerian Pendidikan dan Kebudayaan*, 2019.
- Sholeh, Muh Ibnu, Muh Habibulloh, S Sokip, Asrop Syafi'i, S Sahri, Moh Nashihuddin, Nur 'Azah, Fakhrudin Al Farisy, and S Sulistyorini. “Implementasi Evaluasi Pembelajaran Pendidikan Agama Islam Berbasis Higher Order Thinking Skills (HOTS).” *Al'Ulum Jurnal Pendidikan Islam* 5, no. 1 (2025): 12–26. <https://doi.org/10.54090/alulum.639>.
- Stiggins, Rick. “From Formative Assessment to Assessment FOR Learning: A Path to Success in Standards-Based Schools.” *Phi Delta Kappan* 87, no. 4 (2005): 324–28.
- Subay, Resvia. “Pengembangan Assessment Tes Higher Order Thinking Skills (HOTS) Peserta Didik pada Pembelajaran Matematika Kelas VII Berbasis Model Rasch (Tesis Magister).” *Universitas Negeri Semarang*. Universitas Negeri Semarang, 2020.

- Sudarma, Teguh Febri, Yuni Adha, and Hefi Alberida. "Validity and Reliability of Higher Order Thinking Test Instrument." *Jurnal Pelita Pendidikan* 11, no. 1 (2023): 17–22.
- Sulaiman, Moh. *Sejarah Kebudayaan Islam Kelas XI*. Direktorat KSKK Madrasah Direktorat Jenderal Pendidikan Islam Kementerian Agama RI, 2020.
- Sumintono, Bambang, and Wahyu Widhiarso. *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata, 2015.
- Susilowati, Yayuk, and Sumaji. "Interaksi Bepikir Kritis dengan High Order Thingkig Skill (HOTS) Berdasarkan Taksonomi Bloom." *Jurnal Silogisme* 5, no. 2 (2020): 62–71. <http://journal.umpo.ac.id/index.php/silogisme>.
- Tasrif. "Higher Order Thinking Skills (HOTS) dalam Pembelajaran Social Studies di Sekolah Menengah Atas." *Jurnal Pembangunan Pendidikan: Fondasi dan Aplikasi* 10, no. 1 (2022): 50–61.
- Tennant, Alan, and Ayse A. Küçükdeveci. "Application of the Rasch Measurement Model in Rehabilitation Research and Practice: Early Developments, Current Practice, and Future Challenges." *Frontiers in Rehabilitation Sciences* 4, no. July (2023): 1–17. <https://doi.org/10.3389/fresc.2023.1208670>.
- Tessmer, Martin. *Planning and Conducting Formative Evaluations*. Routledge, 1993. [https://doi.org/https://doi.org/10.4324/9780203061978](https://doi.org/10.4324/9780203061978).
- Treagust, David F. "Development and Use of Diagnostic Tests to Evaluate Students' Misconceptions in Science." *International Journal of Science Education* 10, no. 2 (1988): 159–69. <https://doi.org/10.1080/0950069880100204>.
- Watung, Sjeddie Rianne, and Listriyanti Palangda. "The Influence of Higher Order Thinking Skills (HOTS) and Class Management on Student Learning Outcomes at Tondano 2 State High School." *J-Shelves Of Indragiri (JSI)* 5, no. 1 (2023): 111–24.
- Widana, I Wayan. *Modul Penyusunan Higher Order Thinking Skill (HOTS)*. Direktorat Pembinaan SMA Direktorat Jenderal Pendidikan Dasar dan Menengah Departemen Pendidikan dan Kebudayaan, 2017.
- Wiggins, Grant, and Jay McTighe. *Understanding by Design*. Expanded 2. ASCD, 2005.
- Wijayasari, Estik, Kurniawati Kurniawati, and Murni Winarsih. "Challenge of History Teachers in Teaching and Learning Higher Order Thinking Skills (HOTS)." *Paramita: Historical Studies Journal* 30, no. 1 (2020): 36–45. <https://doi.org/10.15294/paramita.v30i1.20031>.
- Zakariya, Din Muhammad. *Sejarah Peradaban Islam. Sejarah Islam*. Malang: CV. Intrans Publishing, 2018.