



## JURNAL BASICEDU

Volume 7 Nomor 1 Tahun 2023 Halaman 11 - 27

Research & Learning in Elementary Education

<https://jbasic.org/index.php/basicedu>



### Pengembangan Instrumen Tes Berbasis HOTS dengan Pendekatan Pengukuran Rasch pada Pelajaran Matematika Topik Bangun Ruang untuk Siswa Sekolah Dasar

Jumrah<sup>1✉</sup>, Rukli<sup>2</sup>, Sulfasyah<sup>3</sup>

Universitas Muhammadiyah Makassar, Indonesia<sup>1</sup>

Universitas Muhammadiyah Makassar, Indonesia<sup>2,3</sup>

E-mail: [jumrahqeis@gmail.com](mailto:jumrahqeis@gmail.com)<sup>1</sup>, [rukli@unismuh.ac.id](mailto:rukli@unismuh.ac.id)<sup>2</sup>, [sulfasyah@unismuh.ac.id](mailto:sulfasyah@unismuh.ac.id)<sup>3</sup>

#### Abstrak

Tujuan penelitian ini adalah mengembangkan instrumen tes HOTS siswa kelas VI Matematika yang valid dan terpercaya, praktis dan efektif, serta memiliki kualitas yang baik. Instrumen tes ini akan digunakan oleh siswa. Alat uji yang dikembangkan dilakukan penilaian validitas internal dengan bantuan Aiken, serta analisis butir dengan bantuan pengukuran RASCH, dengan tujuan untuk menentukan tingkat kesukaran butir, tingkat kesesuaian butir item, dan deteksi bias. Penelitian ini dilaksanakan di UPT SD Negeri 12 Bontoramba yang terletak di Kabupaten Jeneponto. Dalam penelitian ini, digunakan desain penelitian pengembangan yang diusulkan oleh Martin Tessmer. Rencana Martin membutuhkan penyelesaian empat tahap. Instrumen tes di telaah oleh 3 ahli, uji coba *one to one* diikuti oleh 3 orang siswa, *small group* diikuti oleh 6 orang siswa dan *field tes* diikuti oleh 28 orang siswa. Pengumpulan data terdiri dari informasi kualitatif dan kuantitatif, dan itu dilakukan melalui penggunaan prosedur seperti wawancara, kuesioner, dan pemberian soal tes. Kesimpulan dari penelitian ini keefektifan instrumen tes HOTS yang telah dikembangkan dan menghasilkan *prototype III; (d)* serta tahap *field test (high resistance in revision)*.

**Kata kunci** : pengembangan instrumen tes, HOTS, pengukuran RASCH, ruang bangunan.

#### Abstract

*This study aims to produce valid and reliable, practical and effective student HOTS test instruments in class VI mathematics learning and have good characteristics. The test instruments developed were analyzed for internal validity using Aiken and item analysis using the RASCH measurement to determine the level of difficulty of the items, the level of solubility of the items, and the detection of bias. This research was conducted at UPT SD Negeri 12 Bontoramba, Jeneponto Regency. This study uses the development research design proposed by Martin Tessmer, which consists of four stages. Instrument tests were reviewed by 3 experts, one-to-one trials were attended by 3 students, small group tests were attended by 6 students and field tests were attended by 28 students. Data collection was carried out using interview techniques, questionnaires, and giving test questions so that the data analyzed were qualitative and quantitative data. The conclusion of this study is the effectiveness of the HOTS test instrument that has been developed and produces prototype III; (d) as well as the field test stage (high resistance in revision).*

**Keywords** : development of test instruments, HOTS, RASCH measurements, building space.

Copyright (c) 2023 Jumrah, Rukli, Sulfasyah

✉ Corresponding author :

Email : [jumrahqeis@gmail.com](mailto:jumrahqeis@gmail.com)

DOI : <https://doi.org/10.31004/basicedu.v7i1.4207>

ISSN 2580-3735 (Media Cetak)

ISSN 2580-1147 (Media Online)

Jurnal Basicedu Vol 7 No 1 Tahun 2023  
p-ISSN 2580-3735 e-ISSN 2580-1147

## PENDAHULUAN

Mukhadis dalam Hamzah (2022) mengklaim bahwa abad ke-21 juga dikenal sebagai abad pengetahuan dan bahwa pada era ini segala kemungkinan alternatif pemenuhan kebutuhan hidup dalam berbagai keadaan lebih berbasis pengetahuan. Klaim ini didasarkan pada gagasan bahwa abad ke-21 juga dikenal sebagai abad pengetahuan. Salah satunya adalah inisiatif untuk mengatasi prasyarat untuk pendidikan berbasis pengetahuan (knowledge based education). Menghasilkan sumber daya manusia yang produktif untuk menjawab tantangan kemajuan global dalam berbagai disiplin ilmu merupakan tujuan dari penerapan kurikulum baru dalam sistem pendidikan di Indonesia. Konsekuensinya, kurikulum pendidikan yang saat ini digunakan di Indonesia telah mengalami empat kali revisi dalam kurun waktu tujuh belas tahun. Revisi tersebut dimulai dengan penerapan kurikulum berbasis kompetensi pada tahun 2004, diikuti oleh kurikulum KTSP pada tahun 2006, kurikulum 2013 pada tahun 2013, dan kurikulum terbaru yaitu Kurikulum yang diperkenalkan dan dilaksanakan pada tahun 2015 (Hidayat, 2021), selanjutnya Mukminan dalam Kurniawati & Hadi (2021) Pentingnya penyempurnaan kurikulum pendidikan di Indonesia diartikan untuk menghasilkan lulusan yang memiliki kemampuan berpikir tingkat tinggi, berkepribadian Indonesia, mampu melestarikan budaya bangsa, mampu bersosialisasi budaya, dan berwawasan global. Fitriani, D., Suryana, Y., & Hamdu, G (2018 ) Diungkapkan pula bahwa salah satu ciri lulusan yang diharapkan dapat terbentuk adalah kemampuan berpikir pada tingkat yang lebih tinggi. Rahmah, A. N., & Muharni, L. P. J dalam Rusudianto (2020) Dikatakan bahwa kurikulum 2013 sangat menekankan pada kemampuan berpikir tingkat tinggi (HOTS) siswa. Alhasil, buku yang masuk dalam kurikulum 2013 ini wajib memuat soal-soal yang dapat membantu melatih kemampuan HOTS siswa. Menurut Lewy, L., Zulkardi, Z., & Aisyah, N dalam Nisa & Pahlevi (2021) Ini berfungsi sebagai dasar untuk pemikiran tingkat tinggi dan secara khusus dikenal sebagai taksonomi Bloom. Garis penalaran ini didasarkan pada pengamatan bahwa bentuk-bentuk pembelajaran tertentu membutuhkan tingkat pemrosesan kognitif yang lebih tinggi daripada yang lain, namun menghasilkan keuntungan keseluruhan yang lebih besar. Anderson, W. L., & Krathwohl, R. D (dalam Arifin, 2018) sedangkan taksonomi baru memiliki dua dimensi, yaitu proses kognitif dan pengetahuan, taksonomi Bloom yang asli hanya memiliki satu dimensi, kata penulis argumen tersebut. Mengingat, memahami, menerapkan, menganalisis, mengevaluasi, dan membangkitkan adalah enam subkategori yang membentuk dimensi proses kognitif. Pengetahuan faktual, konseptual, prosedural, dan metakognitif membentuk empat subkategori yang membentuk dimensi pengetahuan (Nurwahida, 2018). Anderson, L. W., & Krathwohl, D. R dalam Lestar (2019) Disebutkan pula bahwa klasifikasi HOTS yang lebih modern tidak lagi hanya menggabungkan satu dimensi (dimensi proses kognitif), melainkan menggabungkan tiga unsur dimensi pengetahuan tertinggi dengan tiga dimensi proses kognitif teratas (menganalisis, mengevaluasi, dan memproduksi). (konseptual, prosedural, dan metakognitif). Ditegaskan pula bahwa klasifikasi HOTS yang lebih modern tidak lagi hanya melibatkan satu dimensi (dimensi proses kognitif), melainkan memadukan tiga unsur dimensi pengetahuan tertinggi dengan tiga unsur dimensi proses kognitif teratas (konseptual, prosedural, dan metakognitif).

Semua pihak kini dapat dengan cepat dan mudah mengakses kekayaan informasi berkat era informasi global modern yang banyak sumbernya tersebar di seluruh dunia (Beddu, 2019). Untuk itu, manusia perlu memiliki kemampuan untuk memperoleh informasi, memilih informasi, mengelola informasi, dan bertindak atas informasi agar berguna dalam kehidupan yang dinamis, kompleks, dan penuh persaingan. Karena tujuan pembelajaran matematika di sekolah adalah untuk membangun keterampilan-keterampilan seperti itu, hal ini dapat dicapai melalui banyak kegiatan pembelajaran matematika (Alfarisa et al., 2021). Aspek kognitif, psikomotorik, dan afektif perilaku siswa yang dibuktikan dengan pergeseran pola pikir siswa yang dominan, baik secara individu maupun kolektif, merupakan komponen yang dievaluasi dan dianalisis. Budiman, A. & Jailani, J dalam Widiyawati (2019) mengisyaratkan bahwa alat evaluasi dapat digunakan dalam bentuk tes maupun non tes. Menurut Docktor dan Heller dalam Masitoh & Aedi (2020) Penggunaan alat penilaian

merupakan komponen penting dari keseluruhan proses evaluasi dalam proses pembelajaran. Penilaian berfungsi sebagai teknik untuk mengevaluasi suatu program pembelajaran, serta kemajuan dan hasil belajar siswa (Desilva et al., 2020).

Penilaian dilakukan secara bertahap, dan nilai peserta ditentukan dengan menjumlahkan nilai yang diperoleh siswa masing-masing untuk setiap tahap. Skor mentah peserta kemudian digunakan untuk memperkirakan tingkat keterampilan mereka (Herman et al., 2021). Tingkat kesulitan setiap langkah tidak diperhitungkan dalam model penskoran seperti itu, yang berarti diperlukan pendekatan alternatif. Salah satu alternatif pendekatan yang dapat digunakan untuk penilaian politomus adalah pendekatan teori respons item. Namun, model penilaian seperti itu tidak selalu tepat. Penskoran PCM adalah model penskoran polytomous, yang berarti menghasilkan lebih dari dua kategori, dan setiap item dalam penilaian mungkin memiliki jumlah kategori respons yang bervariasi (Umami et al., 2021a). Untuk tujuan analisis item dalam penelitian ini, model Rasch digunakan. Di masa lalu, penelitian pengembangan, khususnya di bidang pendidikan matematika, telah menggunakan model Rasch, dilakukan oleh Khumaeroh, S. U., Susongko, P., & Rokhman, M. S (2017) yaitu untuk tahun ajaran 2016–2017, penelitian pembelajaran matematika di SMA Negeri 1 Bojong akan difokuskan pada pengembangan skala sikap siswa terhadap matematika dengan menggunakan model Rasch. Skala ini akan digunakan untuk menunjukkan reliabilitas tes sikap siswa terhadap matematika dengan menggunakan model Rasch ditinjau dari faktor isi, substantif, struktural, eksternal, dan internal. Model Plomp, yang dimodifikasi untuk memenuhi kebutuhan para peneliti, adalah salah satu yang digunakan untuk proses pengembangan sambil menyiapkan ujian standar sikap siswa terhadap matematika. Paket eRm R Programming 3.1.2 dan metode analisis kuantitatif SPSS 17.0 digunakan untuk pengecekan validitas instrumen menggunakan model Rasch. Suryapusparini, B. K., Wardono, W., & Kartono, K (dalam Izzati & Febrian, 2021) menganalisis soal matematika berjenis Higher Order Thinking Skills (HOTS) pada kurikulum 2013 untuk mendukung kemampuan literasi siswa. Penelitian ini dilakukan dengan tujuan untuk menguji soal-soal matematika berjenis Higher Order Thinking Skills (HOTS) pada kurikulum 2013 untuk mendukung kemampuan literasi siswa. Kurikulum 2013 perlu mengembangkan lebih lanjut soal-soal tersebut untuk meningkatkan kemampuan literasi yang lebih baik karena penelitian menunjukkan bahwa soal-soal jenis HOTS yang menuntut kemampuan berpikir tingkat tinggi dapat melatih anak berpikir pada tingkat analisis, penilaian, dan kreativitas. Aplikasi winsteps dapat digunakan untuk melakukan analisis model Rasch, seperti yang dilakukan oleh Soeharto, S., & Rosmayadi, R, Rochman, S., & Hartoyo, Z, dan Adzhar, H., Karim, A. A., & Sahrin, M. U (dalam Izzati & Febrian, 2021).

Isgiyanto, A dalam Lestari (2019) Studi sebelumnya tentang perbandingan skor model Rasch dan model kredit parsial dalam matematika telah dilakukan. Penelitian ini dilakukan dengan menggunakan metodologi retrofitting yang meliputi pemeriksaan butir soal UN Matematika serta data respon butir soal. Partisipan dalam penelitian ini adalah siswa SMP yang berdomisili di daerah Bantul Daerah Istimewa Yogyakarta dan mengikuti ujian nasional matematika SMP tahun pelajaran 2007/2008. 40 soal dan 1016 jawaban soal tingkat soal dari yang mengikuti ujian matematika dijadikan sebagai soal yang diseleksi. Pengambilan sampel menggunakan metode simple random sampling yang berasal dari 12 sekolah dan 1016 orang. Tujuan penelitian adalah untuk memastikan karakteristik item; kapasitas peserta untuk menanggapi item; penilaian model Rasch dan kredit parsial; dan pentingnya fungsi informasi model Rasch dan ujian matematika kredit parsial. Data berupa soal-soal ujian matematika dievaluasi untuk menghasilkan rubrik penskoran, dan peserta yang memilih jawaban yang sesuai diberi skor setinggi mungkin (skor 3). Menurut beratnya kesalahan, tiga pilihan yang salah (*distractors*) ditempatkan ke dalam salah satu dari tiga kategori: skor dua untuk kategori ringan, satu untuk kategori sedang, dan satu untuk kategori berat (skor 1). (skor 0). Berdasarkan beberapa penelitian terdahulu, belum banyak yang berhasil dalam menciptakan materi tes *Higher Order Thinking Skills* (HOTS) untuk pembelajaran matematika yang berbentuk soal esai. Konstruksi tes penilaian HOTS (*Higher Order Thinking Skills*) Matematika, sebaliknya, biasanya hanya memperhatikan satu dimensi, yaitu dimensi

proses berpikir. Akibatnya, sedikit studi telah dilakukan pada produksi penilaian. Riset dilakukan selama seluruh proses pengembangan, sampai pada titik di mana item tersebut dianalisis. Untuk keperluan menganalisis item yang termasuk dalam penelitian ini, program Winsteps versi 3.73 dan model Rasch keduanya digunakan.

Berdasarkan informasi yang disampaikan di atas, maka perlu dibuat instrumen tes HOTS yang dibakukan untuk tujuan penilaian kemampuan berpikir tingkat tinggi yang dimiliki siswa pembelajaran matematika kelas VI. Jika dibandingkan dengan tuntutan kurikulum 2013 yang menuntut siswa untuk berpikir tingkat tinggi, maka jumlah soal HOTS (*higher order thinking skill*) yang masuk dalam buku siswa yang disediakan pemerintah untuk keperluan tes keterampilan berpikir tingkat tinggi siswa sangat rendah, meskipun faktanya pertanyaan-pertanyaan ini termasuk dalam buku. Untuk keperluan penelitian ini akan dibuat instrumen tes HOTS berbasis model Rasch untuk siswa matematika kelas VI. Jumlah bank soal yang dapat digunakan untuk menilai dan mengajarkan HOTS siswa dalam konteks pembelajaran matematika akan semakin banyak karena pengembangan instrumen tes HOTS siswa pada pembelajaran matematika kelas VI berbasis model Rasch. kelompok VI. Oleh karena itu, penulis akan melakukan penelitian dengan judul “Pengembangan Instrumen Tes Berbasis Higher Order Thinking Skills Dengan Pendekatan Model RASCH Pada Pelajaran Matematika Topik Bangun Ruang untuk Siswa Kelas VI UPT SD Negeri 12 Bontoramba. Pada studi ini menggunakan model penelitian *Formative Research Tessmer*. Model pengembangan ini memiliki tahapan-tahapan yang membantu mengembangkan perangkat ke tingkat kesulitan yang valid dan reliabel serta identifikasi yang tepat. Selama tahap evaluasi, peneliti menggunakan lembar validasi untuk memverifikasi keefektifan perangkat yang dikembangkan. Kemudian untuk melihat reliabel, kesulitan serta keunikan bisa dilihat pada uji coba small test. Melalui langkah-langkah di atas, pengembangan peralatan akan memenuhi standar kualitas yang ditetapkan.

## **METODE**

Penelitian ini menggunakan model penelitian pengembangan (Research and Development) dengan tipe *formative research Tessmer* (Sugiyono, 2017). Ada 4 tahapan dalam penelitian ini, yakni tahap pendahuluan (*preliminary*), evaluasi diri (*self-evaluation*) serta evaluasi formatif (*formative evaluation*) atau *prototyping*, mencakup *expert review* (telaah ahli), uji individu (*one-to-one*) dengan resistensi rendah terhadap revisi dan kelompok kecil (*small group*) serta tahap uji lapangan (*field test*) dengan resistensi tinggi terhadap revisi (I.W. Puwardana et al., 2021).

Penelitian ini membutuhkan langkah-langkah atau tahapan untuk memperoleh hasil akhir *prototype* alat uji yang linier dengan target penelitian.

### **Pendahuluan (Preliminary)**

Di tahap ini, peneliti mengkaji berbagai sumber referensi relevan terkait penelitian ini. Informasi dan teori yang sudah dihimpun kemudian dijadikan dasar bagi peneliti untuk menentukan subjek dan tempat uji coba, salah satunya menghubungi kepala sekolah dan guru matematika di sekolah pilihan. Setelah itu, dilakukan persiapan yang lain, meliputi penyusunan jadwal penelitian dan prosedur kolaborasi bersama para guru kelas juga dilakukan di lokasi penelitian (Hanafi et al., 2022).

### **Evaluasi Diri (Self-evaluation)**

Analisis dan desain merupakan bagian dari tahapan ini. Peneliti menggunakan *self-evaluation* untuk meneliti sendiri rancang alat uji HOTS atau keterampilan berpikir tingkat tinggi yang akan dibuat.

### **Analisis Kurikulum**

Kurikulum matematika, literatur, tantangan dan tuntutan masa depan ditelaah pada tahapan ini untuk memperoleh alat uji yang dapat menilai HOTS atau keterampilan berpikir tingkat tinggi.

### Analisis Siswa

Informasi mengenai karakteristik dan jumlah siswa yang sesuai perencanaan pengembangan alat uji dicari dalam tahapan ini. Latar belakang perkembangan dan pengetahuan kognitif siswa (subjek uji coba) merupakan karakteristik dasar yang dicari dalam penelitian ini.

### Analisis Materi

Dalam melakukan identifikasi, perincian, dan merancang materi pembelajaran, analisa materi harus dilakukan secara sistematis berdasarkan analisis kurikulum yang telah dilakukan. Materi-materi inti yang diperlukan sebagai rambu-rambu pengembangan alat uji dapat diidentifikasi menggunakan analisis ini.

### Desain

Kisi-kisi soal pada alat uji, soal-soal alat uji keterampilan berpikir tingkat tinggi, kriteria jawaban soal uji, serta pedoman penilaian dirancang oleh peneliti pada tahapan ini. Rancangan ini kemudian digunakan sebagai *prototype* dan produk dimana ada 3 (tiga) karakteristik yang menjadi fokus dari setiap *prototype*, yaitu: isi, konstruk, serta bahasa, yang dapat diuraikan secara lebih rinci melalui Tabel 1 berikut ini (Siagian & Sagala, 2021).

**Tabel 1. Karakteristik yang Menjadi Fokus *Prototype***

Karakteristik	Uraian
Isi	<ul style="list-style-type: none"> <li>Linieritas antara soal dengan materi pembelajaran di kelas VI UPT SDN 12 Bontoramba</li> <li>Kejelasan maksud soal uji</li> </ul>
Konstruk	Kesesuaian soal dengan teori pendukung dan indikator: <ul style="list-style-type: none"> <li>Soal-soal uji untuk mengukur keterampilan berpikir tingkat tinggi merepresentasikan permasalahan yang ada</li> <li>Memiliki lebih dari satu strategi dan solusi penyelesaian</li> <li>Sesuai tingkatan siswa kelas VI UPT SDN 12 Bontoramba</li> </ul>
Bahasa	<ul style="list-style-type: none"> <li>Soal uji sesuai dengan PUEBI</li> <li>Soal uji tidak ambigu/ bermakna ganda</li> <li>Soal bersifat komutatif, memakai bahasa yang tidak rumit dan mudah dimengerti</li> </ul>

### Validasi, Evaluasi, dan Revisi (*Prototyping*)

Evaluasi terhadap produk yang sudah diciptakan dilakukan pada tahapan ini dan kemudian diujicobakan kepada tiga grup, yakni *expert review*, *one-to-one* dan *small group*. Hasil rancangan pengembangan *prototype* pertama berdasarkan *self-evaluation* secara paralel diberikan kepada para ahli (*expert review*), siswa (*one-to-one*) dan kelompok kecil (*small group*). Ketiga kelompok tersebut memberikan hasil untuk materi perbaikan.

#### Telaah Ahli (*Expert Review*)

Input dan rekomendasi para ahli diperlukan untuk penyempurnaan instrument. Input dan saran tersebut diperoleh melalui *expert review*. Menurut Hermawan (2019:140), produk rancangan, meliputi konstruk, konten, dan bahasa dari setiap *prototype*, akan diobservasi, dinilai dan dievaluasi.

Input dan rekomendasi ahli tersebut akan dipergunakan untuk memperbaiki dan mengembangkan instrumen uji. Semua respon dan rekomendasi dari para ahli (validator) terhadap rancangan tersebut kemudian dicatat pada lembaran validasi untuk materi perbaikan serta memvalidasi rancangan tersebut.

#### Uji Individu (*One-to-one*)

Selanjutnya, peneliti menguji 3 siswa secara individual (*One-to-one*). Ketiganya dimintai untuk mengomentari atau menanggapi instrumen uji yang dikembangkan. Hasil uji individu tersebut kemudian

digunakan untuk mempertimbangkan hasil uji ahli. Kedua hasil pengujian yang telah dilakukan kemudian direvisi untuk menghasilkan produk tipe 1.

#### **Kelompok Kecil (*Small Group*)**

Produk tipe 1 selanjutnya kembali diuji coba ke dalam *Small Group* beranggotakan 6 (enam) siswa non-subjek, dimana 2 (dua) diantaranya memiliki keterampilan berpikir kritis, dua siswa mempunyai keterampilan berpikir tingkat sedang, dan dua lainnya mempunyai keterampilan berpikir tingkat rendah. Produk tipe 2 kemudian dihasilkan dari uji coba ini.

#### **Uji Coba Lapangan (*Field Test*)**

Komentar, rekomendasi serta hasil uji coba produk tipe 2 kemudian digunakan sebagai acuan untuk memperbaiki rancangan produk tipe 2. Hasil perbaikan kemudian diujicobakan pada tahapan *field test*. Produk yang telah direvisi tersebut kemudian diujicobakan untuk subjek uji coba dengan jangkauan lebih luas.

### **HASIL DAN PEMBAHASAN**

Setelah kegiatan yang berkaitan dengan analisis selesai, tahap selanjutnya adalah membuat atau merancang instrumen tes untuk mengukur kemampuan berpikir tingkat tinggi. Instrumen ini dapat mencakup kisi-kisi tes, pertanyaan tes, kriteria respons tes, lembar kerja, dan panduan penilaian.

Langkah awal yang menjadi tanggung jawab peneliti adalah pengembangan tes untuk mengevaluasi keterampilan berpikir tingkat tinggi. Isi yang telah dipelajari digunakan untuk menginformasikan pengembangan soal tes, yang juga diinformasikan oleh indikator yang terdapat pada materi. Peneliti datang dengan sejumlah item tes yang berbeda yang mungkin berdiri di semua indikator materi. Soal-soal yang telah dirancang merupakan soal-soal uraian, dan memiliki karakteristik yang sama dengan soal-soal berpikir tingkat tinggi, yaitu meliputi kategori menganalisis, mengevaluasi, dan memproduksi.

Untuk membantu validator dalam menentukan ketepatan soal-soal yang termasuk dalam ujian kemampuan berpikir tingkat tinggi tentang aspek koneksi, maka peneliti juga membuat kisi-kisi soal dan kriteria jawaban soal. Indikator pencapaian dan domain kognitif dari setiap item digunakan sebagai dasar untuk mengembangkan kisi-kisi tes, atau dirujuk dalam pengembangannya. Lembar kerja disediakan untuk siswa sehingga mereka dapat memiliki akses yang lebih baik ke fasilitas saat mengerjakan soal. Selain itu, peneliti membuat pedoman penilaian yang memudahkan peneliti, guru, dan peneliti lainnya dalam memberikan evaluasi terhadap hasil tes kemampuan berpikir tingkat tinggi yang telah diselesaikan siswa. Hal ini dilakukan agar peneliti dapat lebih memahami data yang mereka kumpulkan. Pada tahap ini, dikembangkan kisi-kisi, 10 soal untuk esai HOTS, lembar kerja, kriteria jawaban, dan petunjuk penilaian. Pada titik proses ini, pertanyaan dari tipe produk I akan dibuat.

Proses mengembangkan soal untuk ujian dianggap oleh para sarjana sebagai langkah yang menghadirkan tantangan terbesar. Hal ini dimaksudkan agar peneliti dapat membuat soal-soal yang dapat secara akurat menilai kemampuan berpikir tingkat tinggi siswa untuk kegiatan khusus ini. Hal ini diperlukan agar instrumen tes yang dihasilkan dapat digunakan dengan baik.

#### **Tahap *Prototyping* (Validasi, Evaluasi, dan Revisi)**

Tahap *prototyping* berusaha untuk menghasilkan pertanyaan tipe II. Pertanyaan produk tipe I yang dibuat sebelumnya akan dinilai dalam tiga kelompok, termasuk tinjauan ahli, satu lawan satu, dan kelompok kecil. Hasil desain produk tipe I yang dibuat berdasarkan evaluasi diri secara bersamaan dipresentasikan kepada ahli (*expert review*), siswa (*one-on-one*), dan kelompok kecil. Tiga set hasil tersebut akan menjadi bahan revisi. Berikut ini membenarkan hasil tahap prototipe:

#### ***Expert Review***

Validasi oleh pakar (*expert review*) dilakukan oleh peneliti dengan membekali validator yang terdiri dari dua orang instruktur Pendidikan Matematika Universitas Muhammadiyah Makassar, dengan alat berupa

kisi-kisi, soal tes, LKS, kriteria jawaban, dan pedoman evaluasi. Validatornya adalah Bpk. Dr. Agustan S M.Pd (validator 1) dan bapak Dr. Rukli, M.Pd, M.Cs (validator 2) dan satu guru Kelas VI UPT SD Negeri 12 Bontoramba yaitu Bapak Jumakka, S.Pd (validator 3). Pada tahap proses validasi ini, validator mengevaluasi lima aspek instrumen yang telah dikembangkan pada prototipe I. Aspek-aspek tersebut relevan dengan instrumen. Setiap faset berkisar dari yang terendah satu hingga tertinggi lima poin yang memungkinkan. Dimana nilai satu berarti tidak valid, 2 kurang valid, 3 cukup valid, 4 berarti valid dan 5 berarti sangat valid. Validator memberikan penilaiannya yang dapat disimpulkan sebagai berikut: produk layak pakai tanpa modifikasi, produk layak pakai dengan revisi, atau produk tidak layak pakai.

**Tabel 2. Penilaian validator**

Validator	Penilaian
Validator 1	Ada sebagian komponen soal yang perlu direvisi
Validator 2	Ada sebagian komponen soal yang perlu direvisi
Validator 3	Ada sebagian komponen soal yang perlu direvisi

Temuan validasi akan disusun dalam kategori menggunakan skala Likert. Soal matematika HOTS khusus ini akan mengalami modifikasi produk untuk memasukkan masukan atau kritik yang diberikan oleh validator ahli. Temuan evaluasi dari para validator dirangkum dalam tabel berikut.

**Tabel 3. Daftar penilaian validator**

No	Validator	Hasil	
		Skor	Kategori
1	Validator 1	3,33	Sangat baik
2	Validator 2	3,41	Sangat baik
3	Guru Kelas VI	3,37	Sangat baik
Rata-rata		3,37 (sangat baik)	

Berdasarkan temuan rekapitulasi evaluasi validator, kriteria sangat baik menghasilkan skor rata-rata 3,4. Tabel 4 di bawah ini memberikan saran pemutakhiran validator untuk instrumen seperti kisi-kisi soal, soal tes, LKS, kriteria jawaban, dan petunjuk penilaian.

**Tabel 4. Saran revisi dari validator**

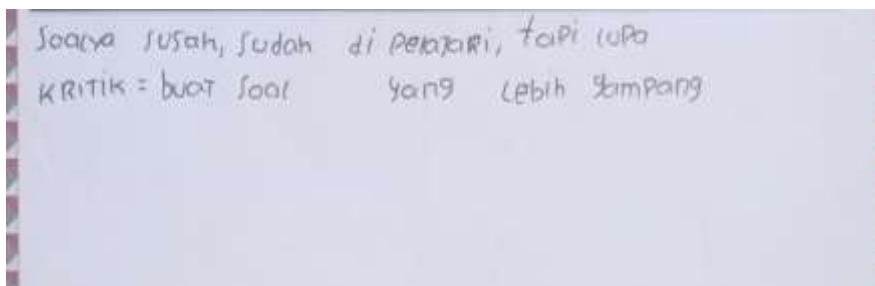
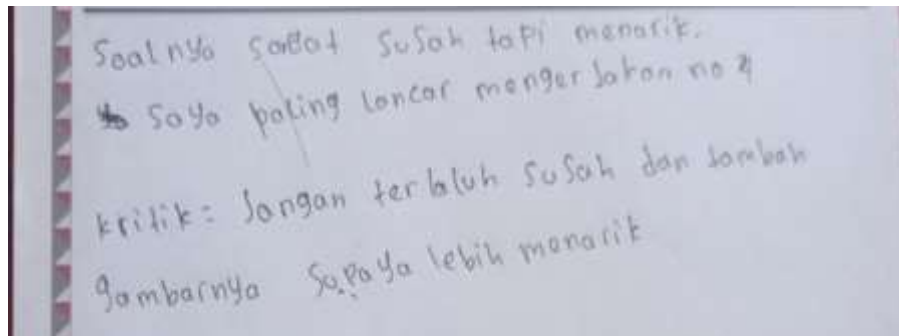
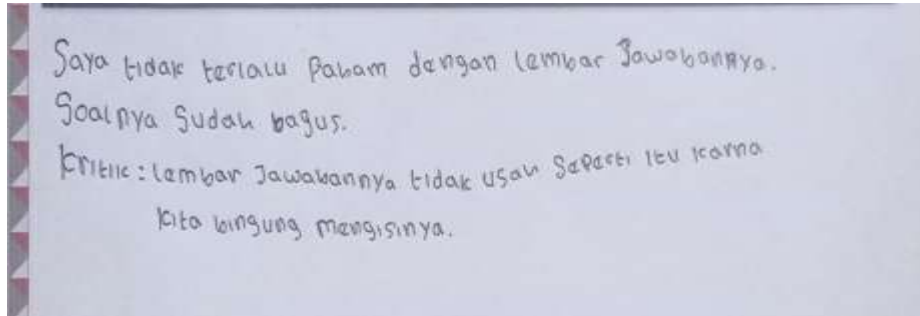
No	Validator	Instrumen	Saran Revisi
1	Validator 1	Soal tes	Ada beberapa konsep yang mesti diperjelas kembali sehingga item tersebut merepresntasikan konsep tersebut sesuai dengan konsep semestinya. Misalnya konsep pesegi panjang atau kubus dari alas sebuah prisma
2	Validator 2	Kisi-kisi tes	Ditambahkan kolom untuk bobot soal
		Soal tes	Petunjuk ditambahkan poin mengerjakan soal yang mudah terlebih dahulu
3	Validator 3	Lembar kerja	Tidak perlu memecah tahapan-tahapan yang terlibat dalam mengatasi masalah karena hal itu akan membuat siswa lebih sulit untuk

	memahaminya.
Kriteria jawaban	Sesuai dengan format lembar kerja

Berdasarkan tabel 4, komentar dan saran yang diberikan oleh validator digunakan sebagai bahan pertimbangan dalam proses perubahan instrumen tes untuk menghasilkan soal produk tipe II dan menilai kemampuan berpikir tingkat tinggi.

a. *One to one*

Selain divalidasi oleh profesional, soal tes keterampilan berpikir tingkat tinggi juga divalidasi oleh tiga orang siswa UPT SD Negeri 12 Bontoramba yang seumurannya dengan subjek penelitian namun tidak menjadi partisipan penelitian. Siswa-siswa ini tidak terlibat dalam penelitian. Berdasarkan temuan peneliti dan saran yang diberikan oleh pengajar di kelas VI UPT SD Negeri 12 Bontoramba, ketiga siswa tersebut terdiri dari satu siswa dengan kemampuan tinggi, satu siswa dengan kemampuan sedang, dan satu siswa dengan kemampuan rendah. Langkah berikutnya melibatkan siswa menawarkan umpan balik mereka pada pertanyaan tes yang telah mereka kerjakan. Pembuatan prototipe II akan diinformasikan temuan jawaban dan komentar siswa, yang akan menjadi bahan pertimbangan.



Gambar 1. Komentar siswa



**b. Small Group**

Hasil produk soal tipe II selanjutnya akan dievaluasi dalam kelompok kecil yang masing-masing terdiri dari enam siswa yang tidak mempelajari mata pelajaran yang sama. Pada tahap prosedur ini, siswa diminta untuk menyelesaikan tes keterampilan berpikir tingkat tinggi yang diperbarui. Hasil percobaan ini menunjukkan bahwa desain tes tidak perlu diubah lebih lanjut sebelum dapat digunakan. Hasil akhir dari tahap ini akan menjadi produk tipe III, yang kemudian akan dievaluasi dengan ukuran sampel yang lebih luas (*field test*).

**c. Field Test**

Pada tahap ini prototipe akan dievaluasi pada subjek penelitian, khususnya kelas VI UPT SD Negeri 12 Bontoramba, setelah sebelumnya dilakukan validasi dan perubahan. Eksperimen khusus di lapangan ini dilakukan pada 25 Juli 2022, dengan total 28 mahasiswa yang dijadikan sebagai subjek penelitian. Tahap ini diselesaikan tanpa menggunakan sumber daya online apa pun. Pada 27 Juli 2022, para peneliti membagikan lembar kerja kepada anak-anak sekaligus memberikan penjelasan tentang apa yang akan mereka lakukan. Sidang berlangsung antara pukul 08.00-10.00 pada tanggal 27 Juli 2022. Setelah siswa menyelesaikan soal-soal tes HOTS, mereka diberi arahan untuk segera mengumpulkan dan menyerahkan kertas jawabannya kepada peneliti. Tingkat kemampuan berpikir tingkat tinggi siswa akan diukur dan ditentukan berdasarkan hasil analisis skor yang diperoleh dari pekerjaan siswa sendiri. Hasil karya siswa juga akan dinilai nilai keterandalan, tingkat kesulitan, dan deteksi bias dari soal yang telah dikembangkan. Selain itu, analisis akan melihat pertanyaan yang telah dikembangkan.

**1. Kevalidan instrumen tes**

Kami menggunakan Uji Validitas Aiken untuk menentukan apakah pertanyaan yang diajukan valid atau tidak. Validitas soal ditentukan dalam dua tahap yaitu: i) Validitas soal HOTS, yang meliputi mengetahui frekuensi indikator HOTS dalam soal yang dikembangkan; dan ii) Validasi soal secara umum yang meliputi mengetahui kaidah penulisan, kesesuaian soal dengan kompetensi dasar yang akan diujikan, kaidah penulisan soal berbentuk esai, dan tata bahasa yang digunakan. Kedua tahapan ini dilakukan untuk memastikan validitas soal. Dua instruktur di jurusan Pendidikan Matematika Unismuh Makassar bertugas mengevaluasi dan memvalidasi setiap langkah. Berikut ringkasan dasar temuan dari uji validasi yang dilakukan pada soal HOTS:

**Tabel 5. Hasil Validasi Instrumen Soal**

Jenis Validasi	Rata-rata Validitas Aiken
Soal Umum	0,79
Soal HOTS	0,76
Total	1,55
Rata-rata	0,775

Berdasarkan temuan yang disajikan pada Tabel 5, diketahui bahwa alat uji yang dirancang mencapai nilai validitas keseluruhan sebesar 0,775 dengan rata-rata kategori valid. Evaluasi yang dilakukan oleh spesialis memberikan skor mulai dari 1 (sangat tidak memuaskan) hingga 4 (sangat memuaskan) (Sangat Baik). Setelah itu, skor dihitung untuk digunakan dalam analisis selanjutnya dengan menggunakan Uji Validitas Aiken untuk menentukan sejauh mana pertanyaan tersebut sah.

**2. Kepraktisan Instrumen Tes**

Hasil angket respon siswa yang diberikan kepada siswa setelah mereka menjawab pertanyaan yang disajikan digunakan untuk menentukan seberapa praktis pertanyaan tersebut dapat diterapkan. Siswa akan memberikan tanggapan terhadap pertanyaan dalam kuesioner yang berkisar dari sangat setuju hingga tidak

setuju hingga setuju hingga sangat setuju hingga tidak setuju hingga sangat tidak setuju, berdasarkan pengalaman pribadi mereka sendiri. Pada tahap *one to-one* dan *small group*, kuesioner dibagikan kepada peserta. I ciri-ciri HOTS berupa kebaruan dan kemenarikan soal (sesuai pengalaman siswa); dan (ii) kepraktisan soal berupa bahasa dan alokasi waktu Tujuannya untuk menentukan kedua hal tersebut.

**Tabel 6. Hasil Kepraktisan Instrumen Tes HOTS**

Jenis tahapan	Rata-rata presentase angket respon siswa
One-to-one	74,67
Small group	80,00

Tabel 6 menunjukkan bahwa pertanyaan-pertanyaan tersebut sudah praktis pada beberapa tingkatan. Hasil skor pada *one-to-one* maupun *small group* yang menunjukkan hasil pada rentang  $63 \leq P < 81$  dengan kriteria praktis/baik menunjukkan hal tersebut.

### 3. Keefektifan Instrumen Tes HOTS

Keberhasilan soal-soal HOTS yang disinggung dalam penelitian ini bergantung pada ketepatan pengukuran kemampuan berpikir tingkat tinggi siswa. Agar soal-soal HOTS dianggap efektif dalam penelitian ini, syarat-syarat berikut harus dipenuhi: Analisis spesialis menunjukkan bahwa soal-soal tersebut berguna untuk menentukan apakah siswa memiliki keterampilan berpikir tingkat tinggi atau tidak. Menampilkan hasil yang sesuai dengan yang diantisipasi. Menurut temuan penyelidikan ini, kemampuan berpikir tingkat tinggi siswa rata-rata mencapai tingkat tertinggi dari apa yang dianggap cukup. Karena siswa biasanya tidak terbiasa menjawab pertanyaan berdasarkan HOTS, masuk akal untuk mengasumsikan bahwa keterampilan berpikir tingkat tinggi mereka termasuk dalam kategori kurang atau cukup. Kriteria kecukupan ditentukan berdasarkan kenyataan yang ada di lapangan (Sagala & Andriani, 2019).

Pada tahap *expert review* dilakukan penelitian tentang keefektifan yang didasarkan pada pendapat para ahli. Pakar atau validator akan menggunakan pengetahuan dan pengalamannya untuk mengevaluasi (memberikan tanggapan) apakah soal-soal yang telah disusun bermanfaat atau tidak dalam menguji kemampuan berpikir tingkat tinggi siswa. Menurut individu yang berpengalaman, keefektifan soal mendapat skor 81,66% pada kategori yang dikhususkan untuk mengevaluasi keefektifan. Selain itu, berdasarkan hasil inkuiri kemampuan berpikir tingkat tinggi siswa diketahui bahwa nilai rata-rata kemampuan tersebut termasuk dalam kategori cukup yaitu 5,3. Hal ini menunjukkan keefektifan pertanyaan penelitian karena hasilnya sesuai dengan harapan. Menurut Akker dalam Sagala & Andriani (2019), sebuah pertanyaan dianggap efektif jika para ahli setuju dengan keefektifannya dan menghasilkan hasil yang diinginkan. Ini konsisten dengan gagasan bahwa mengajukan pertanyaan yang tepat akan mengarah pada hasil yang diinginkan. Selain itu, hasil tes kemampuan berpikir tingkat tinggi ini menunjukkan kemungkinan bahwa pertanyaan dapat memengaruhi kemampuan siswa untuk mengembangkan kemampuan berpikir tingkat tinggi.

Nilai akhir yang diperoleh setelah mengerjakan soal tes kemampuan berpikir tingkat tinggi menentukan data yang tampak pada hasil tes yang digunakan untuk mengukur kemampuan berpikir tingkat tinggi siswa. Setelah data hasil tes diperiksa, data tersebut diubah menjadi data kualitatif sehingga dapat diketahui tingkat kemampuan berpikir tingkat tinggi yang dimiliki siswa. Temuan investigasi tingkat kemampuan berpikir tingkat tinggi siswa ditampilkan pada tabel 4.3 di bawah halaman.

**Tabel 7. Analisis data hasil tes HOTS**

Nilai siswa	frekuensi	Presentase (%)	Kategori
$80 < \text{nilai} \leq 100$	0	0	Sangat baik
$60 < \text{nilai} \leq 80$	8	28,57	Baik
$40 < \text{nilai} \leq 60$	14	50,00	Cukup
$20 < \text{nilai} \leq 40$	6	21,43	Kurang
$0 \leq \text{nilai} \leq 20$	0	0	Sangat kurang
Jumlah Subjek	28	100	
Rata-rata Nilai		53,57	Cukup

Dari total 28 subjek uji coba, analisis data menunjukkan bahwa tidak ada siswa yang termasuk dalam kategori sangat baik, 8 siswa (28,57%) termasuk dalam kategori baik, 14 siswa (50,00%) dalam kategori sedang, 6 siswa (21,43%) termasuk dalam kategori kurang, dan tidak ada siswa yang termasuk dalam kategori sangat kurang. Temuan ini didasarkan pada kenyataan bahwa tidak ada siswa yang termasuk dalam kategori sangat kurang.

#### 4. Hasil Analisis menggunakan model RASCH

Model RASCH kemudian digunakan bersama dengan perangkat lunak Winsteps untuk melakukan analisis terhadap data yang diperoleh dari hasil pengujian. Keluaran dari software Winsteps digunakan untuk menentukan sejumlah parameter item yang kompatibel dengan model Rasch. Selain itu, uji reliabilitas item secara keseluruhan dilakukan, dan nilai alpha Cronbach dihitung sebagai hasilnya. Sementara itu, batasan item yang dinyatakan fit dengan model ditunjukkan oleh Outfit MNSQ, Outfit ZSTD, dan nilai korelasi item dengan pertanyaan secara keseluruhan. Yaitu, apabila nilai Outfit MNSQ berada diantara 0,5 sampai dengan 1,5; nilai Outfit ZSTD berada diantara -2,0 sampai dengan 2,0; serta nilai korelasi butir dengan skor total berada diantara 0,4 sampai dengan 0,85 (Sumintono & Widhiarso, 2015, p. 12).

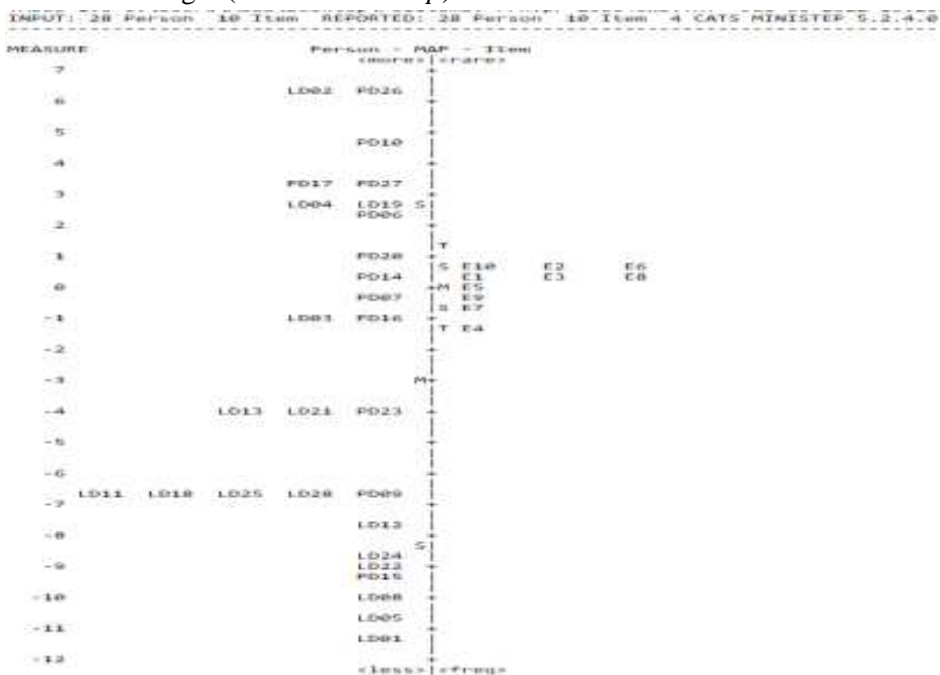
Berdasarkan analisa data menggunakan software Winsteps, hasilnya disajikan secara lengkap pada tabel 8

**Tabel 8. Ringkasan Statistik**

	Keterangan	Nilai
Logit	Person	-2.91
	Item	0.00
Reliabilitas	Person reliability	0.95
	Item reliability	0.17
	Alpha Cronbach	0,96
Outfit MNSQ	Person	0,84
	Item	0,84
Outfit ZSTD	Person	-0,04
	Item	-0,14

Kolom person dan item pada Tabel 8 sama-sama menampilkan nilai Outfit Mean Squared (Outfit MNSQ) sebesar 0,84. Kriteria kecocokan mencakup nilai 0,84 yang berada dalam kisaran  $0,5 < \text{MNSQ} < \text{ZSTD} < 2,0$ , yang menunjukkan bahwa data mungkin memiliki nilai rasional. Alhasil, item atau item tersebut dapat dimanfaatkan sebagai alat uji materi geometri karena pada umumnya mengikuti model Rasch.

1. Analisis Peta Wright (Person-Item Map) Data Politomi



Gambar 2. Analisis Peta Wright

Keluaran Winsteps, yaitu peta Wright, menggambarkan pentingnya kemahiran siswa dalam mengatasi masalah. Bakat tertinggi terdapat pada anak dengan kode LD02 dan PD26, sedangkan kemampuan terendah terdapat pada individu dengan kode LD11, LD18, LD25, LD28, dan PD09..

2. Analisis Butir Soal

Beberapa analisis butir, antara lain tingkat kesulitan butir, tingkat kesesuaian butir, dan pendeteksian bias, dapat dilakukan untuk mengetahui kualitas butir uraian ini.

a. Analisis tingkat kesulitan soal

Person: REAL SEP.: 4.18 REL.: .95 ... Item: REAL SEP.: .45 REL.: .17

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S. E.	INFIT   MNSQ	ZSTD	OUTFIT   MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OB5%	MATCH EXP%	Item
2	58	28	.55	.52	1.20	.70	.98	.19	.87	.83	78.6	82.5	E2
6	58	28	.55	.52	1.84	2.27	1.65	1.04	.78	.83	64.3	82.5	E6
10	58	28	.55	.52	.45	-2.14	.30	-1.29	.86	.83	92.9	82.5	E10
1	59	28	.28	.52	.57	-1.52	.39	-.95	.87	.83	92.9	82.7	E1
3	59	28	.28	.52	.96	-.02	.98	.21	.79	.83	85.7	82.7	E3
8	59	28	.28	.52	1.02	.17	.76	-.16	.83	.83	78.6	82.7	E8
5	60	28	.00	.52	1.22	.75	1.13	.41	.83	.83	78.6	82.5	E5
9	61	28	-.27	.52	.52	-1.82	.32	-1.11	.88	.83	89.3	82.2	E9
7	63	28	-.82	.53	1.24	.87	.99	.22	.85	.83	75.0	81.6	E7
4	65	28	-1.38	.53	.98	.04	.89	.08	.78	.84	82.1	81.5	E4
MEAN	60.0	28.0	.00	.52	1.00	-.07	.84	-.14			81.8	82.3	
P.S.D	2.2	.0	.62	.00	.40	1.31	.40	.71			8.4	.4	

Gambar 3. Hasil analisis item measure

Nilai logit (ukuran) yang tinggi menunjukkan tingkat kesulitan yang tinggi untuk item tersebut. Panduan diberikan oleh Suminto dan Widhiarso (2015) untuk mengklasifikasikan objek-objek tersebut ke dalam empat kategori:

1. Nilai measure < -1 = item sangat mudah
2. Nilai measure -1 s.d. 0 = item mudah
3. Nilai measure 0 s.d. 1 = item sulit
4. Nilai measure > 1 = item sangat sulit

b. Analisis tingkat kesesuaian soal (item fit)

Tingkat kesesuaian item digunakan untuk menilai seberapa akurat item tersebut sesuai dengan model. Saat melakukan pengukuran, item fit menunjukkan apakah item beroperasi secara normal atau tidak. Jika ada elemen yang tidak sesuai, subjek mungkin belum memahami pertanyaan dengan benar. Analisis temuan item fit tercantum di bawah ini.

Person: REAL SEP.: 4.18 REL.: .95 ... Item: REAL SEP.: .45 REL.: .17

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASURE-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Item
6	58	28	.55	.52	1.84	2.27	1.65	1.04	A .78	.83	64.3	82.5	E6
7	63	28	-.82	.53	1.24	.87	.99	.22	B .85	.83	75.0	81.6	E7
5	60	28	.00	.52	1.22	.75	1.13	.41	C .83	.83	78.6	82.5	E5
2	58	28	.55	.52	1.20	.70	.98	.19	D .87	.83	78.6	82.5	E2
8	59	28	.28	.52	1.02	.17	.76	-.16	E .83	.83	78.6	82.7	E8
3	59	28	.28	.52	.96	-.02	.98	.21	e .79	.83	85.7	82.7	E3
4	65	28	-1.38	.53	.98	.04	.89	.08	d .78	.84	82.1	81.5	E4
1	59	28	.28	.52	.57	-1.52	.39	-.95	c .87	.83	92.9	82.7	E1
9	61	28	-.27	.52	.52	-1.82	.32	-1.11	b .88	.83	89.3	82.2	E9
10	58	28	.55	.52	.45	-2.14	.30	-1.29	a .86	.83	92.9	82.5	E10
MEAN	60.0	28.0	.00	.52	1.00	-.07	.84	-.14			81.8	82.3	
P.SD	2.2	.0	.62	.00	.40	1.31	.40	.71			8.4	.4	

Gambar 4. Data hasil item fit order

Menurut Boone, Staver, & Yale (2014), nilai *outfit means-square*, *outfit z-standard*, dan *point measure correlation* adalah kriteria yang digunakan untuk melihat tingkat kesesuaian butir. Panduan untuk menilai kriteria kesesuaian butir menurut Boone, et al (2014) adalah sebagai berikut

1. Nilai Outfit Mean Square (MNSQ) yang diterima :  $0,5 < MNSQ < 1,5$
2. Nilai outfit Z-standard (ZSTD) yang diterima:  $-2,0 < ZSTD < +2,0$
3. Nilai Point Measure Correlation yang diterima:  $0,4 < pt \text{ measure corr} < 0,85$

Alagumalai, Curtis, dan Hungi (2005) mengklasifikasikan nilai *Point Measure Correlation* sebagai sangat baik (>0,40), baik (0,30-0,39), cukup (0,20-0,29), tidak mampu membedakan (0,00-0,19), dan membutuhkan pemeriksaan item (<0,00) karena korelasi ukuran titik secara teoritis setara dengan korelasi titik-biserial dalam teori tes klasik.

c. Analisis adanya bias (DIF)

DIF class/group specification is: DIF=\$S1W1

Person CLASS/	Obs-Exp	DIF	DIF	Person CLASS/	Obs-Exp	DIF	DIF	DIF	JOINT CONTRAST	Rasch-Welch t	Mantel Chi-squ	Size CUMULOR	Active Slices	Item Number	Name	
L	-.03	.51	.69	P	.04	-.02	.78	-.53	1.04	.51	24 .6155			4	1 E1	
L	-.12	1.43	.67	P	.16	-.60	.76	2.03	1.01	2.01	24 .0560	.6667	.4142	4	2 E2	
L	.03	.02	.71	P	-.04	.60	.80	-.58	1.07	-.54	24 .5937			4	3 E3	
L	.09	-2.20	.80	P	-.11	-.60	.76	-1.60	1.10	-1.46	25 .1569			4	4 E4	
L	.00	.02	.71	P	.00	-.02	.78	.04	1.05	.04	24 .9685	1.0000	.3173	4	5 E5	
L	.00	.51	.69	P	-.01	-.60	.80	-.09	1.06	-.09	23 .9297	.6667	.4142	4	6 E6	
L	.02	-1.03	.74	P	-.03	-.60	.76	-.43	1.06	-.41	25 .6865	.0000	1.000	.00	4	7 E7
L	-.09	.98	.68	P	.12	-.60	.76	1.58	1.02	1.56	24 .1329	1.0000	.3173	4	8 E8	
L	.03	-.49	.73	P	-.04	-.02	.78	-.48	1.06	-.45	24 .6576			4	9 E9	
L	.07	.02	.71	P	-.09	1.25	.81	-1.23	1.08	-1.14	23 .2657			4	10 E10	

Gambar 5. Tabel ringkasan DIF

Berdasarkan tabel di atas terlihat bahwa setiap butir soal tidak mengandung satu pun DIF, dimana nilai probabilitasnya semua butir soal diatas 5% (0,05).

**d. Analisis instrumen**

Hasil analisis instrument secara keseluruhan ditampilkan sebagai berikut

SUMMARY OF 28 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	21.4	10.0	-2.91	1.05	.91	.02	.84	-.04	
SEM	1.1	.0	1.05	.12	.08	.14	.09	.14	
P. SD	5.9	.0	5.44	.60	.43	.74	.49	.75	
S. SD	6.0	.0	5.54	.61	.44	.76	.50	.76	
MAX.	32.0	10.0	6.33	2.75	2.23	1.67	2.26	1.34	
MIN.	11.0	10.0	-11.43	.66	.90	-1.35	.60	-1.32	
REAL RMSE	1.26	TRUE SD	5.29	SEPARATION	4.18	Person RELIABILITY	.95		
MODEL RMSE	1.23	TRUE SD	5.30	SEPARATION	4.30	Person RELIABILITY	.95		
S.E. OF Person MEAN = 1.05									
Person RAW SCORE-TO-MEASURE CORRELATION = .98 (approximate due to missing data)									
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .96 SEM = 1.20 (approximate due to missing data)									
STANDARDIZED (50 ITEM) RELIABILITY = .99									
SUMMARY OF 10 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	60.0	28.0	.00	.52	1.00	-.07	.84	-.14	
SEM	.7	.0	.21	.00	.13	.44	.13	.24	
P. SD	2.2	.0	.62	.00	.40	1.31	.40	.71	
S. SD	2.4	.0	.65	.00	.42	1.30	.42	.75	
MAX.	65.0	28.0	.55	.53	1.84	2.27	1.65	1.64	
MIN.	58.0	28.0	-1.30	.52	.45	-2.14	.30	-1.29	
REAL RMSE	.56	TRUE SD	.25	SEPARATION	.45	Item RELIABILITY	.17		
MODEL RMSE	.52	TRUE SD	.32	SEPARATION	.62	Item RELIABILITY	.28		
S.E. OF Item MEAN = .21									

**Gambar 6. Analisis instrumen**

Reliabilitas item 0,17, reliabilitas orang 0,95, dan Alpha Cronbach 0,96 pada tabel di bawah ini. Setiap siswa diharapkan mampu berpikir pada tingkat yang setinggi-tingginya, menjadikan pengembangan keterampilan berpikir tingkat tinggi sebagai tujuan akhir. Oleh karena itu, diperlukan informasi tentang derajat kemampuan berpikir tingkat tinggi setiap siswa sebagai langkah awal dalam upaya peningkatan kemampuan berpikir.

(Nur Aini & Sulistyani, 2019).

**Pembahasan**

Istiyono, E., Mardapi, D., & Suparno, S dalam Hanifah (2019) mengatakan bahwa penilaian dapat dilakukan baik secara lisan maupun tertulis. Tes tertulis digunakan untuk melakukan penilaian tertulis. Selain menggunakan asesmen berupa tes untuk mengajarkan kemampuan berpikir siswa dan berdampak pada mengidentifikasi kemampuan berpikir siswa, Suhaesti J. berpendapat bahwa melakukan asesmen merupakan salah satu teknik untuk mengetahui apakah siswa memiliki kemampuan berpikir tingkat tinggi. Tandilling, E dalam Hanik & Ngazizah (2021) mencatat bahwa anak-anak mengalami kesulitan memahami matematika karena metode tanya jawab dan latihan tanya jawab guru yang kurang beragam di kelas. Lebih lanjut Sukriadi, Kartono, & Wiyanto (2021) berpendapat bahwa belajar matematika harus melibatkan interaksi sosial yang aktif dan pemikiran reflektif, termasuk bertanya, membela, mengeksplorasi, membuat sketsa, membangun, menggunakan, menjelaskan, mengembangkan, dan membuktikan. Budiman, A., & Jailani, J dalam Yulianto & Widodo (2020) mengatakan bahwa selain belum adanya alat asesmen yang khusus dibuat untuk melatih HOTS atau kemampuan berpikir tingkat tinggi siswa, ketidakmampuan guru dalam menyusun instrumen evaluasi HOTS menjadi kendala yang mereka hadapi. Kemampuan sains anak Indonesia masih tergolong rendah, terbukti dari hasil survey TIMSS, dan salah satu faktor penyebabnya adalah siswa di Indonesia kurang terlatih dalam menyelesaikan soal-soal yang mengukur HOTS, dan penilaian HOTS masih kurang. dan alat penilaian masih kurang. Ditemukan juga bahwa bakat ilmiah anak-anak Indonesia masih cukup rendah,

masalah masih muncul di sekolah, dan pertanyaan seringkali mengevaluasi keterampilan yang berkaitan dengan ingatan daripada berpikir tingkat tinggi (Kuswari, 2020).

Wirda, M. A., Berutu, N., Rahmad, R., & Rohani, R dalam Yulianto & Widodo (2020) Ditemukan bahwa di antara bentuk-bentuk tes tersebut ada yang telah dibakukan, artinya telah melalui pengujian yang ketat untuk validitas dan reliabilitasnya serta kelayakan aplikasi praktisnya. Selain itu Puspendik Balitbang Depdiknas dalam Umami (2021) menjelaskan bahwa tes standar adalah tes yang pertanyaannya telah mengalami analisis subyektif dan statistik. Jenis pengujian yang baru-baru ini mendapatkan popularitas adalah pengujian standar. Selain itu dinyatakan bahwa tes yang dibakukan memenuhi kriteria tes yang baik, terutama kriteria validitas, reliabilitas, dan objektivitas. Tes standar juga dapat digunakan untuk waktu yang cukup lama dan dapat diterapkan pada berbagai objek yang mencakup area yang luas, selain fakta bahwa tes tersebut telah dikategorikan berdasarkan usia dan tingkat kelas. Ujian berbentuk esai biasanya dievaluasi sebagian berdasarkan langkah-langkah yang harus dilakukan untuk menjawab butir-butir soal dengan benar.

Penilaian dilakukan sebagai upaya untuk mengukur tingkat ketercapaian indikator pembelajaran dan untuk mengumpulkan informasi tentang berbagai kemajuan belajar yang telah terjadi pada siswa.. Sejalan dengan penelitian relevan yang dilakukan oleh Yulianto & Widodo, (2020) Temuan penelitian menunjukkan bahwa ada perbedaan yang dapat ditemukan antara item dalam hal tingkat tantangan dan penerapannya. Semua item dalam analisis memenuhi ketiga prasyarat tersebut, yang menunjukkan bahwa item-item tersebut memiliki tingkat kecocokan yang baik dan terkendali. Sebagai konsekuensinya, kumpulan soal tes berbasis HOTS termasuk dalam kategori sangat baik karena mampu membedakan berbagai derajat kemampuan berpikir tingkat tinggi yang dimiliki siswa.

Dari nilai tersebut dapat disimpulkan bahwa tingkat konsistensi jawaban siswa cukup tinggi, dan kualitas butir soal dalam instrumen tes yang digunakan memiliki reliabilitas yang cukup baik yaitu 0,17. baik yaitu 0,96. baik yaitu 0,96. baik yaitu 0,96. baik yaitu 0,96. baik yaitu 0,96. baik yaitu 0,96. Bagus.

## **KESIMPULAN**

Pada subjek konstruksi spasial, penelitian telah dilakukan dengan tujuan mengembangkan instrumen pengujian kemampuan berpikir tingkat tinggi untuk digunakan dengan populasi siswa. Dari hasil dan pembahasan tahapan pengembangan yang telah dilakukan dapat diambil kesimpulan bahwa proses pengembangan instrumen tes berbasis HOTS menggunakan model pengembangan *tipe/formative research Tessmer* yang terdiri dari empat tahap, antara lain (a) pendahuluan, dimana peneliti mencari referensi mengenai instrumen tes yang akan dikembangkan; (b) tahap evaluasi diri, peneliti merancang instrumen keterampilan berpikir tingkat tinggi berdasarkan analisis yang telah dilakukan pada tahap sebelumnya; (c) evaluasi formatif (*prototyping*), yang meliputi tinjauan ahli dan satu-ke-satu (ketahanan rendah terhadap revisi), di mana instrumen tes diuji; dan (d) kesimpulan. Hasil pengujian pada tahap ini digunakan untuk menganalisis efisiensi instrumen tes HOTS yang telah dibuat dan menghasilkan prototipe III; (d) serta tahap uji lapangan (resistensi tinggi dalam revisi), khususnya uji coba lapangan di UPT SDN 12 Bontoramba kelas VI untuk mengukur tingkat kemampuan berpikir tingkat tinggi siswa, ketergantungan, tingkat kesulitan, dan deteksi bias pada instrumen tes yang telah dibuat.terdiri dari 4 tahapan yaitu: (a) *preliminary*, pada tahap ini peneliti mencari referensi tentang instrumen tes yang ingin dikembangkan; (b) tahap *self evaluation*, pada tahap ini peneliti merancang instrumen kemampuan berpikir tingkat tinggi sesuai dengan analisis yang telah dilakukan pada tahap sebelumnya; (c) tahap *formative evaluation (prototyping)* yang meliputi *expert reviews* dan *one-to-one (low resistance to revision)*, pada tahap ini instrument tes divalidasi oleh *expert reviews* yang terdiri dari 3 validator diantaranya 2 dosen Pendidikan matematika dan seorang guru Kelas VI UPT SDN 12 Bontoramba, serta uji coba *one-to-one* kepada 3 orang siswa untuk dimintai komentar mengenai tes berpikir tingkat tinggi yang telah dikembangkan yang menghasilkan *prototype* II, dan *small group*, merupakan tahapan



uji coba *prototype* I kepada 6 orang siswa. Hasil uji coba pada pada tahap ini, digunakan untuk menganalisis keefektifan instrumen tes HOTS yang telah dikembangkan dan menghasilkan *prototype* III; (d) serta tahap *field test (high resistance in revision)*, yaitu uji coba lapangan di UPT SDN 12 Bontoramba pada kelas VI untuk mengetahui tingkat kemampuan berpikir tingkat tinggi siswa, reliabilitas, tingkat kesukaran, dan deteksi adanya adanya bias instrumen tes yang telah dikembangkan.

## DAFTAR PUSTAKA

- Alfarisa, F., Supriadi, S., Susilawati, S., Rahimah, A. D., & Yunia, Y. (2021). Pengembangan Instrumen Higher Order Thingking Skill (Hots) Matematika Untuk Siswa Sekolah Dasar. *Proseding Didaktis: Seminar Nasional Pendidikan Dasar*, 6(1), 279–290.
- Arifin, Z. (2018). Pengembangan Instrumen Analisis Kesalahan Untuk Mengukur Critical Thinking Skill Berdasarkan Kompetensi Inti Kurikulum 2013. *Eduma: Mathematics Education Learning And Teaching*, 7(1), 53–62. <https://doi.org/10.24235/Eduma.V7i1.2734>
- Beddu, S. (2019). Implementasi Pembelajaran Higher Order Thinking Skills (Hots) Terhadap Hasil Belajar Peserta Didik. *Jurnal Pemikiran Dan Pengembangan Pembelajaran*, 1(3), 71–84.
- Desilva, D., Sakti, I., & Medriati, R. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Fisika Berorientasi Hots (Higher Order Thinking Skills) Pada Materi elastisitas Dan Hukum Hooke. *Jurnal Kumparan Fisika*, 3(1), 41–50. <https://doi.org/10.33369/Jkf.3.1.41-50>
- Desiriah, E., & Setyarsih, W. (2021). Tinjauan Literatur Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi (Hots) Fisika Di Sma. *Orbita: Jurnal Kajian, Inovasi Dan Aplikasi Pendidikan Fisika*, 7(1), 79. <https://doi.org/10.31764/Orbita.V7i1.4436>
- Hamzah, H., Hamzah, M. I., & Zulkifli, H. (2022). Systematic Literature Review On The Elements Of Metacognition-Based Higher Order Thinking Skills (Hots) Teaching And Learning Modules. *Sustainability*, 14(2), 813. <https://doi.org/10.3390/Su14020813>
- Hanafi, M., Syamsuri, S., & Mutaqin, A. (2022). Pengembangan Instrumen Soal Higher Order Thinking Skills (Hots) Matematika Berdasarkan Brookhart Konteks Motif Batik Pandegelang Pada Siswa Mts. *Media Pendidikan Matematika*, 10(1), 43. <https://doi.org/10.33394/Mpm.V10i1.5207>
- Hanifah, N. (2019). Pengembangan Instrumen Penilaian Higher Order Thinking Skill (Hots) Di Sekolah Dasar. *Current Research In Education: Conference Series Journal*, 1(1), 5.
- Hanik, A., & Ngazizah, N. (2021). Pengembangan Instrumen Tes Berbasis Higher Order Thinking Skills (Hots) Kelas V Madrasah Ibtidaiyah. *Jurnal Pendidikan Dasar*, 2(1), 74–84. <https://doi.org/10.37729/Jpd.V2i1.967>
- Herman, H., Rahim, A. R., & Syamsuri, A. S. (2021). Analisis Instrumen Tes Hasil Belajar Berbasis Higher Order Thinking Skill (Hots). *Jurnal Riset Dan Inovasi Pembelajaran*, 1(3), 88–101. <https://doi.org/10.51574/Jrip.V1i3.65>
- Hidayat, A. A. (2021). *Menyusun Instrumen Penelitian & Uji Validitas-Reliabilitas*. Health Books Publishing.
- I.W. Puwardana, Sariyasa, & I.N. Suastika. (2021). Pengembangan Instrumen Penilaian Berbasis Hots Pada Materi Pengolahan Data Dalam Kehidupan Sehari-Hari Untuk Siswa Kelas Vi Sd. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, 11(2), 147–156. <https://doi.org/10.23887/Jpepi.V11i2.625>
- Izzati, N., & Febrian, F. (2021). Kemampuan Mahasiswa Calon Guru Dalam Mengembangkan Instrumen Tes Hasil Belajar Kategori Higher Order Thinking Skill. *Jurnal Gantang*, 6(1), 75–82. <https://doi.org/10.31629/Jg.V6i1.2672>
- Khumaeroh, S. U., Susongko, P., & Rokhman, M. S. (2017). Penyusunan Skala Sikap Peserta Didik Terhadap Matematika Dengan Penerapan Model Rasch. *Jpmp (Jurnal Pendidikan Mipa Pancasakti)*, 1(1). <https://doi.org/10.24905/Jpmp.V1i1.787>



- 27 *Pengembangan Instrumen Tes Berbasis HOTS dengan Pendekatan Pengukuran Rasch pada Pelajaran Matematika Topik Bangun Ruang untuk Siswa Sekolah Dasar – Jumrah, Rukli, Sulfasyah*  
DOI : <https://doi.org/10.31004/basicedu.v7i1.4207>
- Kurniawati, R. P., & Hadi, F. R. (2021). Pelatihan Pengembangan Instrumen Evaluasi Berbasis Hots Untuk Guru Sekolah Dasar. *Jurnal Altifani Penelitian Dan Pengabdian Kepada Masyarakat*, 1(4), 267–276. <https://doi.org/10.25008/Altifani.V1i4.182>
- Kuswari, R. I. (2020). *Pengembangan Lks Berbasis Higher Order Thinking Skill (Hots) Dalam Meningkatkan Hasil Belajar Matematika Kelas Iv Di Min 3 Tulungagung Dan Sdi Qurrota A'yun Ngunut Tulungagung*. Iain Tulungagung.
- Lestari, S. A. P. (2019). Pengembangan Instrumen Asesmen Higher Order Thinking Skill (Hots) Pada Materi Himpunan Kelas Vii Smp. *Jkpm (Jurnal Kajian Pendidikan Matematika)*, 4(2), 111. <https://doi.org/10.30998/Jkpm.V4i2.3862>
- Masitoh, L. F., & Aedi, W. G. (2020). Pengembangan Instrumen Asesmen Higher Order Thinking Skills (Hots) Matematika Di Smp Kelas Vii. *Jurnal Cendekia : Jurnal Pendidikan Matematika*, 4(2), 886–897. <https://doi.org/10.31004/Cendekia.V4i2.328>
- Ningsih, I. I. S., Sugiarno, S., & Hamdani, H. (N.D.). Pengembangan Instrumen Tes Untuk Mengukur Kemampuan Higher Order Thinking (Hot) Dikaji Dari Literasi Matematis. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, 9(12). <https://doi.org/http://dx.doi.org/10.26418/Jppk.V9i12.44086>
- Nisa, S., & Pahlevi, T. (2021). Pengembangan Instrument Penilaian Hots Berbantuan Quizizz Pada Mata Pelajaran Kearsipan Smk. *Edukatif: Jurnal Ilmu Pendidikan*, 3(5), 2146–2159. <https://doi.org/10.31004/Edukatif.V3i5.756>
- Nur Aini, D. F., & Sulistyani, N. (2019). Pengembangan Instrumen Penilaian E-Quiz (Electronic Quiz) Matematika Berbasis Hots (Higher Of Order Thinking Skills) Untuk Kelas V Sekolah Dasar. *Edumaspul: Jurnal Pendidikan*, 3(2), 1–10. <https://doi.org/10.33487/Edumaspul.V3i2.137>
- Nurwahida, O. (2018). *Pengaruh Pendekatan Higher Order Thinking Skills (Hots) Terhadap Hasil Belajar Ilmu Pengetahuan Sosial Murid Kelas Iv Sd Inpres Bontomanai Kecamatan Tamalate Kota Makassar*.
- Rusdianto, A. R., Susanta, A. S., & Muktadir, A. M. (2020). Pengembangan Instrumen Penilaian Berbasis Higher Order Thinking (Hot) Pelajaran Matematika Kelas Iv Sekolah Dasar. *Jurnal Pembelajaran Dan Pengajaran Pendidikan Dasar*, 3(1), 1–19. <https://doi.org/10.33369/Dikdas.V3i1.12110>
- Siagian, Q. A., & Sagala, P. N. (2021). Development Of Test Instruments To Measure High Order Thinking Skill (Hots) Mathematics Of Students In Mts 2 State Of Medan City. *Mathline : Jurnal Matematika Dan Pendidikan Matematika*, 6(2), 154–174. <https://doi.org/10.31943/Mathline.V6i2.222>
- Sugiyono. (2017). *Metode Penelitian Pendidikan*. Bandung: Alfabeta,Cv.
- Umami, R., Rusdi, M., & Kamid, K. (2021a). Pengembangan Instrumen Tes Untuk Mengukur Higher Order Thinking Skills (Hots) Berorientasi Programme For International Student Assessment (Pisa) Pada Peserta Didik. *Jp3m (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika)*, 7(1), 57–68.
- Umami, R., Rusdi, M., & Kamid, K. (2021b). Pengembangan Instrumen Tes Untuk Mengukur Higher Order Thinking Skills (Hots) Berorientasi Programme For International Student Assessment (Pisa) Pada Peserta Didik. *Jp3m (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika)*, 7(1), 57–68. <https://doi.org/10.37058/Jp3m.V7i1.2069>
- Widiyawati, Y., Nurwahidah, I., & Sari, D. S. (2019). Pengembangan Instrumen Integrated Science Test Tipe Pilihan Ganda Beralasan Untuk Mengukur Hots Peserta Didik. *Saintifika*, 21(2), 1–14.
- Yulianto, A., & Widodo, A. (2020). Disclosure Of Difficulty Distribution Of Hots-Based Test Questions Through Rasch Modeling. *Indonesian Journal Of Primary Education*, 4(2), 197–203. <https://doi.org/10.17509/Ijpe.V4i2.29318>