**PENGEMBANGAN PERANGKAT TES *HIGHER THINKING SKILLS* UNTUK MATERI STATISTIKA SMP*****DEVELOPMENT OF A TEST OF HIGHER THINKING SKILLS IN STATISTICS AT JUNIOR HIGH SCHOOL***Anggit Prabowo^{*1}, Tatang Herman², Siti Fatimah³, Muh. Faiz Setyono⁴^{1,4} Universitas Ahmad Dahlan, Tamanan, Kec. Banguntapan, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55191^{1,2,3} Universitas Pendidikan Indonesia, Jl. Dr. Setiabudi No.229, Isola, Kec. Sukasari, Kota Bandung, Jawa Barat 40154¹anggit.prabowo@pmat.uad.ac.id, ²tatangherman@upi.edu, ³sitifatimah@upi.edu, ⁴mfsetyono2@gmail.com**Corresponding Author*

Abstrak: Tujuan penelitian ini adalah untuk: (1) mengembangkan tes untuk mengukur *higher thinking skills* siswa kelas VIII SMP/MTs pada materi statistika dan (2) mengetahui karakteristik instrumen tes tersebut. Dengan proses pengembangan analisis, desain, pengembangan, implementasi, dan evaluasi, penelitian ini merupakan studi pengembangan dengan paradigma pengembangan ADDIE. Tes yang dikembangkan terdiri atas 10 butir soal pilihan ganda dan 5 butir soal uraian yang telah divalidasi oleh ahli. Teknik analisis data dilakukan secara kuantitatif dan deskriptif. Subjek uji coba yaitu 30 siswa kelas VIII SMP Negeri 1 Berbah. Berdasarkan hasil uji coba menunjukkan rata-rata indeks kesukaran butir-butir soal pilihan ganda adalah 0,44 (*moderate*) sedangkan rata-rata indeks kesukaran soal uraian adalah 0,35 (*moderate*). Rata-rata daya beda butir soal pilihan ganda adalah 0,40 (*acceptable*) dan daya beda untuk soal uraian adalah 0,47 (*acceptable*). Koefisien reliabilitas instrumen tes soal pilihan ganda dan uraian berturut-turut adalah 0,85 dan 0,74 dengan kriteria sangat tinggi. Dengan demikian, tes yang dikembangkan layak untuk digunakan untuk mengukur *higher thinking skills* siswa kelas VIII SMP/MTs pada materi statistika.

Kata Kunci: tes, *higher thinking skills*, statistika

Abstract: The aims of this research were: (1) to develop a test to measure the higher thinking skills of Grade VIII students of SMP/MTs in statistics and (2) to find out the characteristics of the test. With analysis, design, development, implementation, and evaluation of development processes, this research is a development study employing the ADDIE development paradigm. The developed test consists of 10 multiple choice and 5 essay items that have been validated by experts. Data analysis techniques were carried out quantitatively and descriptively. The test subjects were 30 students of class VIII SMP Negeri 1 Berbah. Based on the test results, the average difficulty index of multiple-choice items was 0.44 (*moderate*) while for essay items was 0.35 (*moderate*). The average differential index of the multiple-choice items was 0.40 (*acceptable*) and for essay items was 0.47 (*acceptable*). The reliability coefficient of the multiple choice and essays was 0.85 and 0.74 respectively with very high criteria. The coefficient for essay questions was 0.74 with high criteria. Thus, the test developed is feasible to be used to measure the higher thinking skills of Grade VIII students of SMP/MTs in statistics material.

Keywords: test, higher order thinking skills, statistics

Cara Sitasi: Prabowo, A., Herman, T., Fatimah, S., & Setyono, M. F. (2023). Pengembangan perangkat tes higher thinking skills untuk materi statistika SMP. *Math Didactic: Jurnal Pendidikan Matematika*, 9(3), 400-409.

<https://doi.org/10.33654/math.v9i3.1163>

Matematika adalah mata pelajaran yang mendasari perkembangan berbagai ilmu pengetahuan. Penguasaan materi matematika yang baik akan berdampak positif pada penguasaan materi mata pelajaran lainnya (Shim et al., 2017). Karenanya, matematika diajarkan pada siswa di berbagai level di Indonesia. Konten materi matematika yang dipelajari di setiap jenjang pendidikan di Indonesia khususnya pendidikan dasar dan menengah didasarkan pada capaian pembelajaran yang tertuang dalam kurikulum terbaru, Kurikulum Merdeka.

Kemampuan siswa Indonesia di bidang matematika dapat diketahui dari berbagai studi yang dilaksanakan di level nasional maupun internasional. Salah satunya adalah *Programme for International Student Assessment* (PISA). Tujuan PISA adalah untuk mengukur keterampilan dan kompetensi siswa dalam membaca, matematika, dan sains dalam konteks dunia nyata pada usia 15 tahun (OECD, 2019). Indonesia mengikuti studi PISA sejak tahun 2001. Di tahun 2018, hasil PISA siswa Indonesia menunjukkan rata-rata kemampuan matematika siswa Indonesia adalah 379 (OECD, 2019). Sementara itu, rata-rata kemampuan seluruh siswa peserta adalah 489. Hal ini menunjukkan bahwa rata-rata kemampuan matematika siswa Indonesia berada di bawah rata-rata peserta dari negara-negara peserta studi PISA.

Tes PISA menggunakan soal yang berbeda dari yang sering ditemui siswa. Karakter soalnya memerlukan kemampuan penalaran, pemecahan masalah, kreativitas, dan bersifat kontekstual (Wardhani & Rumiati, 2011). Karakteristik soal seperti itu merupakan karakteristik dari soal untuk mengukur keterampilan berpikir tingkat tinggi (Widana, 2017) atau *higher order thinking skills* (HOTS). HOTS merupakan jenis berpikir yang non algoritmik dan kompleks (Barak & Judy, 2019). HOTS sering dikaitkan dengan taksonomi tujuan pendidikan oleh Bloom (Scully, 2017). Level berpikir tingkat tinggi berada pada tahapan analisis, sintesis, dan evaluasi (Zohar & Dori, 2003). Untuk taksonomi Bloom revisi, level HOTS berada pada tahap analisis, evaluasi, dan kreasi (Schraw & Robinson, 2011). Faktanya, siswa Indonesia jarang berlatih menyelesaikan soal-soal yang menuntut pemecahan masalah kontekstual (Wijaya et al., 2014), sebagai salah satu karakteristik soal HOTS. Akibatnya, siswa Indonesia masih memiliki keterampilan berpikir tingkat tinggi yang masih terbatas (Putri & Zulkardi, 2018).

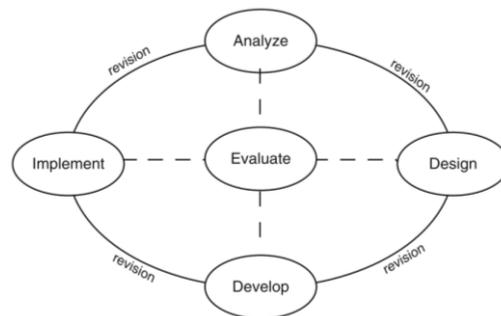
Rendahnya HOTS siswa disebabkan oleh berbagai faktor. Salah satu penyebabnya adalah sumber belajar yang memuat soal tersebut masih terbatas ketersediaannya di sekolah dan di toko buku (Wijaya et al., 2015). Selain itu, pemahaman guru tentang konsep HOTS dan aktivitas pengukuran HOTS siswa masih dalam kategori kurang (Retnawati et al., 2018). Untuk meningkatkan HOTS, siswa harus dibiasakan untuk menyelesaikan soal-soal HOTS (Arifin & Retnawati, 2017). Apabila hal ini tidak dilaksanakan, potensi siswa dalam menyelesaikan HOTS tidak optimal.

Memperhatikan kondisi tersebut, perlu adanya pembiasaan siswa dalam menyelesaikan soal-soal HOTS. Upaya ini tentu harus didukung dengan ketersediaan soal-soal yang memadai untuk berbagai materi matematika. Penelitian ini mengembangkan soal-soal HOTS untuk materi statistika SMP. Pertimbangan pemilihan materi ini adalah karena dari hasil laporan-laporan ujian nasional menunjukkan bahwa penguasaan materi statistika siswa SMP di Indonesia masih rendah (Prabowo, Anggoro, & Rahmawati, 2018). Salah satu buktinya adalah penelitian Thirafi di tahun 2017 menunjukkan bahwa persentase penguasaan materi statistika siswa Indonesia hanya 48,6% (Sriwahyuni & Maryati, 2022). Penelitian ini dibatasi pengembangan soal statistika untuk kompetensi

dasar: menganalisis data berdasarkan distribusi data, nilai rata-rata, median, modus, dan sebaran data untuk mengambil kesimpulan, keputusan dan membuat prediksi.

Metode Penelitian

Penelitian ini menggunakan konsep pengembangan ADDIE sebagai metode penelitian pengembangan. Tahap-tahap model ini meliputi *analyze*, *design*, *develop*, *implement*, dan *evaluate* seperti terlihat pada Gambar 1. Model ADDIE dipilih karena berfungsi sebagai kerangka untuk memandu situasi yang kompleks dan cocok untuk mengembangkan produk pendidikan dan sumber belajar lainnya, dan karena itu merupakan salah satu metode yang paling efektif (Branch, 2009).



Gambar 1. Konsep Pengembangan ADDIE

Penelitian ini dimulai dengan melakukan analisis tentang perlunya pengembangan tes tipe HOTS untuk siswa SMP pada materi statistika. Pengembangan ini diperlukan karena kondisi di lapangan menunjukkan kemampuan siswa pada HOTS masih rendah dan ketersediaan referensi soal tipe HOTS masih terbatas. Pada tahap desain, peneliti merancang tes yang hendak dikembangkan dengan menyusun kisi-kisi soal dengan mengacu pada kompetensi yang hendak diukur. Tes yang dikembangkan terdiri atas 10 soal berformat pilihan ganda yang dilengkapi dengan 4 pilihan jawaban dan 5 soal uraian. Pada tahap ini peneliti juga mendesain lembar validasi yang akan digunakan untuk melakukan validasi tes oleh ahli. Pada tahap pengembangan, peneliti menulis butir soal dan melakukan validasi isi (*content validity*) kepada dua orang ahli di bidang penilaian pembelajaran matematika. Tes yang sudah divalidasi selanjutnya diimplementasikan pada 30 siswa kelas IX di SMPN 1 Berbah Yogyakarta. Hasil implementasi selanjutnya dianalisis untuk mendapatkan informasi karakteristik butir soal yang meliputi indeks kesukaran, daya beda, dan reliabilitas. Kegiatan evaluasi dilakukan pada setiap fase pengembangan. Data hasil uji coba selanjutnya dianalisis kemudian digunakan sebagai dasar melakukan revisi sehingga menghasilkan tes yang baik.

Hasil Penelitian dan Pembahasan

Hasil

Hasil analisis kondisi di lapangan berdasarkan berbagai studi menunjukkan bahwa keterampilan berpikir tingkat tinggi siswa di Indonesia pada materi matematika masih rendah. Salah satu penyebabnya adalah siswa tidak terbiasa dilatih dan berlatih untuk menyelesaikan soal-soal

bertipe HOTS. Ketersediaan soal-soal HOTS di sekolah masih terbatas. Dengan demikian, pengembangan tes tipe HOTS untuk materi matematika perlu dikembangkan untuk memenuhi kebutuhan di lapangan. Mengingat materi matematika sangat luas, pengembangan tes pada penelitian ini dibatasi pada materi statistika dengan kompetensi seperti telah dijelaskan di awal.

Pada tahap desain, peneliti menyusun kisi-kisi dari kompetensi yang hendak diukur. Kisi-kisi memuat tujuan pengembangan tes, indikator penilaian, jenis tes, dan banyaknya butir tes. Tujuan pengembangan tes adalah untuk mengembangkan tes untuk mengukur keterampilan berpikir tingkat tinggi siswa SMP pada materi statistika dengan kompetensi dasar: menganalisis data berdasarkan distribusi data, nilai rata-rata, median, modus, dan sebaran data untuk mengambil kesimpulan, keputusan dan membuat prediksi. Tes yang didesain terdiri atas 10 butir soal pilihan ganda dan 5 butir soal uraian. Masing-masing butir soal pilihan ganda memiliki 4 pilihan jawaban. [Tabel 1](#) merupakan kisi-kisi tes yang dikembangkan.

Tabel 1. Kisi-kisi Tes

Kompetensi Dasar	Indikator Soal	Bentuk Soal	Level Kognitif	Banyak butir	
Menganalisis data berdasarkan distribusi data, nilai rata-rata, median, modus, dan sebaran data untuk mengambil kesimpulan, keputusan dan membuat prediksi	Disajikan sebuah diagram lingkaran, siswa dapat menyelesaikan masalah sebaran data.	PG	C5	2	
	Disajikan sebuah data, siswa dapat menentukan modus.	PG	C5	1	
	Disajikan sebuah data, siswa dapat membuat kesimpulan.	PG	C6	1	
	Disajikan sebuah masalah kontekstual, siswa dapat menentukan rata-rata dan mengambil kesimpulan.		PG	C4, C5	3
				C6	2
		Uraian	C4, C5	4	
	Disajikan sebuah diagram batang, siswa dapat membuat prediksi.	PG	C5	1	
Disajikan masalah kontekstual, siswa dapat menentukan median	Uraian	C5	1		

Dari kisi-kisi yang sudah didesain, selanjutnya dilakukan pengembangan dengan melakukan penulisan butir-butir soal dengan mengacu pada aturan-aturan menulis butir-butir soal pilihan ganda (Haladyna, T. M. & Downing, 1989) dan uraian. Soal-soal tersebut kemudian divalidasi oleh dua orang ahli di bidang penilaian pembelajaran matematika. Adapun kedua ahli tersebut adalah pendidik di perguruan tinggi yang memiliki pengalaman mengajar dan melakukan penilaian lebih dari 7 tahun. Dalam penilaian ini validator memberikan penilaian terkait dengan validitas isi tes yang dikembangkan yang mencakup kesesuaian butir tes dengan materi, konstruksi, dan bahasa. Hasil penilaian validator tersaji pada [Tabel 2](#).

Tabel 2. Validasi Ahli

Komponen Penilaian	Pilihan Ganda	Uraian
Kesesuaian materi	Sangat baik	Sangat baik
Konstruksi	Sangat baik	Sangat baik
Bahasa	Sangat baik	Sangat baik

Dari [Tabel 2](#) terlihat bahwa baik butir soal pilihan ganda maupun uraian semuanya telah memenuhi kesesuaian materi, konstruksi, dan bahasa. Dengan demikian secara isi, tes yang dikembangkan telah memenuhi validitas isi.

Butir-butir soal yang valid selanjutnya dirakit menjadi sebuah instrumen tes dan diujicobakan kepada 30 siswa kelas VII di SMPN 1 Berbah. Hasil uji coba berikutnya dilakukan analisis terhadap tingkat kesukaran, daya pembeda, dan reliabilitasnya menggunakan bantuan paket program Microsoft Excel dan SPSS yang hasilnya tersaji pada [Tabel 3](#).

Tabel 3. Rata-rata Indeks Kesukaran, Daya Beda, dan Koefisien Reliabilitas Tes

Kriteria	Pilihan Ganda	Uraian
Indeks Kesukaran	0,44	0,35
Daya Beda	0,40	0,47
Reliabilitas	0,85	0,74

Berdasar [Tabel 3](#) diperoleh informasi bahwa rata-rata indeks kesukaran butir soal pilihan ganda adalah 0,44 (*moderate*) dan butir soal uraian adalah 0,35 (*moderate*). Rata-rata daya beda butir soal pilihan ganda adalah 0,40 (*acceptable*) dan butir soal uraian adalah 0,47 (*acceptable*). Koefisien reliabilitas instrumen tes soal pilihan ganda dan uraian berturut-turut adalah 0,85 dan 0,74 dengan kriteria sangat tinggi.

Pembahasan

Setiap individu penting untuk memiliki keterampilan berpikir tingkat tinggi. Individu yang memiliki keterampilan ini akan terbiasa untuk berpikir pada level yang lebih tinggi, sehingga mereka mampu untuk membuat keputusan dari permasalahan-permasalahan yang kompleks (Saraswati & Agustika, 2020). Seiring perkembangan zaman dengan permasalahan yang semakin beragam, keterampilan berpikir tingkat tinggi menjadi kompetensi yang sangat dibutuhkan pada masa sekarang dan era mendatang (Winarso, 2014). Mengingat pentingnya peran keterampilan berpikir tingkat tinggi, maka pembelajaran di sekolah hendaknya melatih keterampilan ini kepada seluruh siswa sesuai dengan level berpikirnya.

Pengembangan tes berorientasi HOTS adalah salah satu upaya untuk melatih keterampilan berpikir tingkat tinggi. Melalui ketersediaan soal-soal HOTS diharapkan mampu menyediakan sumber belajar bagi siswa untuk melatih keterampilan berpikir tingkat tinggi. Karena faktanya, keterampilan ini masih jarang dilatih kepada siswa di sekolah (Prasetyani et al., 2016) karena tidak tersedianya referensi yang kurang memadai (Wijaya et al., 2015). Ketersediaan sumber referensi memberikan tambahan sumber belajar dan berlatih siswa dalam menyelesaikan soal-soal yang melatih keterampilan berpikir tingkat tinggi mereka. Dampaknya diharapkan pada akhirnya siswa mampu meningkatkan keterampilan mereka dalam berpikir secara kompleks, memecahkan masalah, dan menemukan penyelesaian-penyelesaian baru berdasar pengetahuan yang mereka miliki. Dengan memberikan tes yang mengukur keterampilan berpikir tingkat tinggi, mampu meningkatkan keterampilan berpikir tingkat tinggi seorang individu (Mohamed & Lebar, 2017).

Penelitian ini menghasilkan seperangkat instrumen tes untuk melatih dan mengukur keterampilan berpikir tingkat tinggi siswa SMP pada materi statistika. Tes yang dikembangkan terdiri atas soal-soal pilihan ganda (10 butir) dan soal uraian (5 butir). Sejatinya, terdapat berbagai format tes yang mampu digunakan. Namun, dengan pertimbangan agar dapat digunakan pada responden yang banyak dengan penskoran yang mudah, maka salah satu format tes yang dalam pengembangan ini adalah jenis tes pilihan ganda. Tes jenis ini mampu mengukur kemampuan berpikir tingkat tinggi (Clay, 2001) serta memberikan pilihan jawaban yang beragam (Prabowo et al., 2021). Tes ini juga mengukur tingkat pemikiran kompleks, seperti pemikiran kritis dan keterampilan memecahkan masalah (Haladyna et al., 2019). Penambahan butir soal uraian diharapkan mampu memberikan informasi yang lebih mendalam kepada guru terhadap jawaban dan level berpikir siswa. Penambahan

butir ini juga sebagai pemenuhan salah satu syarat bahwa untuk mengukur keterampilan berpikir tingkat tinggi diharapkan menggunakan bentuk soal yang bervariasi (Widana, 2017).

Tes dikembangkan dengan mengacu pada langkah-langkah pengembangan butir soal untuk menghasilkan butir tes yang baik. Kualitas hasil tes akan sangat tergantung pada kualitas soal (Haladyna, 1992). Dalam mengembangkan tes, menulis soal merupakan langkah yang fundamental (Haladyna & Downing, 2009). Dalam mengembangkan butir soal, butir-butir ditelaah dan direvisi berulang kali oleh ahli materi, ahli tes, dan penelaah yang peka sebelum diujikan (Potenza & Stocking, 1997). Pada pengembangan tes ini, dua orang ahli telah menelaah dan memberikan masukan terhadap tes. Butir-butir tes didesain dengan mengacu pada 43 taksonomi penulisan butir soal (Haladyna, T. M. & Downing, 1989). Konstruksi butir soal juga mengacu pada teknik penyusunan butir soal untuk mengukur keterampilan berpikir tingkat tinggi dengan cara membalik butir soal, menggunakan pengecoh berkualitas, dan mengetuk 'multiple neurons' (Scully, 2017).

Tes yang dikembangkan dalam penelitian ini telah memenuhi kriteria validitas. Jenis validitas yang ditekankan pada pengembangan ini adalah validitas isi. Validitas isi adalah sarana kualitatif untuk memastikan bahwa indikator-indikator itu menyentuh makna suatu konsep seperti yang didefinisikan oleh peneliti (Drost, 2011). Jenis validitas ini ditunjukkan dengan justifikasi ahli terkait dengan kesesuaian butir soal dengan konstruk yang hendak diukur (Christensen et al., 2015). Pendekatan validitas ini dilakukan secara kualitatif melalui justifikasi. Ahli memberikan penilaian dan masukan secara deskriptif terhadap butir-butir soal. Untuk pengembangan selanjutnya dapat diperkuat dengan kajian analisis validitas model lainnya seperti penggunaan teknik statistika.

Selain dinyatakan valid oleh justifikasi ahli, tes yang dikembangkan memiliki karakteristik butir soal yang baik. Dari hasil uji coba diperoleh rata-rata indeks kesukaran butir soal pilihan ganda adalah 0,40 sedangkan untuk soal uraian memiliki indeks kesukaran 0,35. Indeks kesukaran adalah proporsi peserta tes menjawab butir soal dengan benar (Sayyah et al., 2012). Rata-rata indeks kesukaran kedua jenis tes berada pada rentang 0,3 sampai dengan 0,7. Pada interval ini adalah rentang indeks kesukaran yang direkomendasikan (Musa et al., 2018). Pada rentang ini, soal berada pada tingkat kesulitan sedang, yaitu tidak terlalu sulit dan tidak terlalu mudah. Soal ini idealnya harus dijawab dengan benar oleh peserta tes berprestasi tinggi dan salah oleh peserta tes berprestasi rendah. Dengan kata lain, soal-soal ini dapat digunakan untuk membedakan siswa yang berprestasi tinggi dan rendah.

Kemampuan tes dalam membedakan karakteristik siswa berdasar kemampuannya disebut dengan daya beda tes. Daya beda adalah selisih antara persentase peserta ujian berkemampuan tinggi dan peserta berkemampuan rendah yang mendapatkan soal dengan benar (Dhakne-Palwe et al., 2015). Rata-rata daya beda tes butir soal pilihan ganda adalah 0,85 sedangkan rata-rata daya beda soal uraian adalah 0,74. Sementara itu, daya beda yang ideal adalah lebih dari 0,29 (Shete et al., 2015). Dengan demikian tes yang dikembangkan memiliki daya beda yang baik. Baiknya daya beda pada tes ini selaras dengan baiknya indeks kesukaran. Hal ini sesuai dengan studi korelasi di antara keduanya, bahwa butir soal dengan daya beda baik cenderung memiliki tingkat kesukaran yang baik atau moderat (Sim & Rasiyah, 2006).

Tes yang baik adalah tes yang mampu menghasilkan hasil pengukuran yang konsisten. Artinya, ketika peserta tes yang sama mengerjakan tes yang sama pada waktu yang berbeda, dengan catatan tanpa diberi perlakuan, hasil tesnya akan konsisten. Kekonsistenan tersebut disebut dengan

reliabilitas yang dinyatakan dalam sebuah koefisien. Terdapat berbagai metode untuk mengestimasi reliabilitas suatu tes. Pada studi ini, pengestimasian reliabilitas tes yang dikembangkan menggunakan formula *Cronbach Alpha*. Formula ini merupakan ukuran konsistensi internal yang paling umum dalam ilmu sosial, perilaku, dan pendidikan dan biasanya ditafsirkan sebagai rata-rata dari semua kemungkinan koefisien *split-half* (Mohajan, 2017). Koefisien reliabilitas butir soal pilihan ganda dan uraian pada tes ini adalah 0,85 dan 0,74. Kriteria ini memenuhi syarat koefisien reliabilitas yang baik, karena tes dinyatakan reliabel jika koefisien reliabilitasnya melebihi 0,5 (Gugiu & Gugiu, 2018). Dengan demikian, tes yang telah dikembangkan dalam penelitian ini mampu menghasilkan hasil pengukuran yang konsisten.

Simpulan dan Saran

Simpulan

Penelitian ini mengembangkan perangkat tes untuk melatih dan mengukur keterampilan berpikir tingkat tinggi siswa SMP pada materi statistika pada kompetensi menganalisis data yang berdasar distribusi data, rerata, nilai tengah, modus, dan sebaran data untuk membuat kesimpulan, keputusan, dan prediksi. Tahapan pengembangan yang ditempuh meliputi tahapan *analyze, design, develop, implement, dan evaluate*. Tes terdiri atas terdiri atas 10 butir soal pilihan ganda dan 5 butir soal uraian. Keseluruhannya telah dinyatakan valid secara isi melalui justifikasi ahli. Rata-rata indeks kesukaran butir-butir soal pilihan ganda adalah 0,44 (*moderate*) sedangkan rata-rata indeks kesukaran soal uraian adalah 0,35 (*moderate*). Rata-rata daya beda butir soal pilihan ganda adalah 0,40 (*acceptable*) dan daya beda untuk soal uraian adalah 0,47 (*acceptable*). Koefisien reliabilitas instrumen tes soal pilihan ganda dan uraian berturut-turut adalah 0,85 dan 0,74 dengan kriteria sangat tinggi. Dengan demikian tes ini layak untuk digunakan dalam melatih dan mengukur keterampilan berpikir tingkat tinggi siswa. Dari penelitian ini terbuka penelitian pengembangan untuk materi matematika yang lain mengingat secara umum keterampilan berpikir tingkat tinggi pada materi matematika siswa masih rendah. Analisis validitas dan estimasi reliabilitas dengan menggunakan metode yang lain juga dapat menjadi bahan kajian penelitian berikutnya termasuk penambahan subjek uji coba.

Saran

Tes yang dikembangkan dalam penelitian pengembangan ini dapat dimanfaatkan oleh guru matematika untuk menambah referensi ketersediaan soal berbasis HOTS pada materi statistika. Guru dapat mengadopsi langkah-langkah pengembangan untuk mengembangkan secara mandiri materi-materi matematika yang lainnya. Guru juga dapat mengembangkan dengan melakukan variasi bentuk tes yang lainnya. Pada teknis pelaksanaan tes, guru diharapkan mengondisikan peserta tes sehingga mampu mendapatkan hasil pengukuran yang akurat.

Daftar Pustaka

- Arifin, Z., & Retnawati, H. (2017). Pengembangan instrumen pengukur higher order thinking skills matematika siswa SMA kelas X. *PYTHAGORAS: Jurnal Pendidikan Matematika*, 12(1), 98. <https://doi.org/10.21831/pg.v12i1.14058>
- Barak, M., & Judy, Y. (2019). Enhancing Higher Order Thinking Skills Among Inservice Science Teachers Via Embedded Assessment. *Journal of Science Teacher Education*, 20(5), 459–474.
- Branch, R. M. (2009). *Approach, Instructional Design: The ADDIE*. Springer.
- Christensen, L. B., Johnson, R. B., & Turner, L. A. (2015). *Research Methods, Design, and Analysis*. Pearson.
- Clay, B. (2001). *Is This a Trick Question ? Is This a Trick*. Kansas Curriculum Center.
- Dhakne-Palwe, S., Gujarathi, A., & Almale, B. (2015). Item Analysis of MCQs and Correlation between Difficulty Index, Discrimination Index and Distractor Efficiency in a Formative Examination in Community Medicine. *Journal of Research in Medical Education & Ethics*, 5(3), 254–259. <https://doi.org/10.5958/2231-6728.2015.00052.9>
- Drost, E. A. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives*, 38(1), 105–123.
- Gugiu, C., & Gugiu, M. (2018). Determining the Minimum Reliability Standard Based on a Decision Criterion. *Journal of Experimental Education*, 86(3), 458–472. <https://doi.org/10.1080/00220973.2017.1315712>
- Haladyna, T. M. (1992). *The Effectiveness of Several Multiple-Choice Formats*. 5(1), 73–88. <https://doi.org/10.1207/s15324818ame0501>
- Haladyna, T. M. & Downing, S. M. (1989). A Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education*, 2(1), 37–50.
- Haladyna, T. M., & Downing, S. M. (2009). A Taxonomy of Multiple-Choice Item- Writing Rules. *Applied Measurement in Education*, 2(1), 37–50. <https://doi.org/10.1207/s15324818ame0201>
- Haladyna, T. M., Rodriguez, M. C., & Stevens, C. (2019). Are Multiple-choice Items Too Fat? *Applied Measurement in Education*, 32(4), 350–364. <https://doi.org/10.1080/08957347.2019.1660348>
- Mohajan, H. K. (2017). Two Criteria for Good Measurements in Research: Validity and Reliability. *Annals of Spiru Haret University. Economic Series*, 17(4), 59–82. <https://doi.org/10.26458/1746>
- Mohamed, R., & Lebar, O. (2017). *Authentic Assessment in Assessing Higher Order Thinking Skills*. 7(2), 466–476. <https://doi.org/10.6007/IJARBS/v7-i2/2021>
- Musa, A., Shaheen, S., & Elmardi, A. (2018). Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University Abstract : *Khartoum Medical Journal*, 11(2), 1477–1468.

- OECD. (2019). PISA 2018 Results Combined Executive Summaries. In *OECD: Vol. I,II, II*. <https://doi.org/10.1787/g222d18af-en>
- Potenza, M. T., & Stocking, M. L. (1997). Flawed Items in Computerized Adaptive Testing. *Journal of Educational Measurement*, 34(1), 79–96.
- Prabowo, A., Puspa, P., & Setyawan, F. (2021). Development of test instruments to measure high level thinking ability of two linear variable equations system. *International Journal on Education Insight*, 1(2), 91. <https://doi.org/10.12928/ije.v1i2.3540>
- Prasetyani, E., Hartono, Y., & Susanti, E. (2016). Kemampuan Berpikir Tingkat Tinggi Siswa Kelas Xi Dalam Pembelajaran Trigonometri Berbasis Masalah Di Sma Negeri 18 Palembang. *Jurnal Gantang*, 1(1), 34–44. <https://doi.org/10.31629/jg.v1i1.4>
- Putri, R. I. I., & Zulkardi, Z. (2018). Higher-order thinking skill problem on data representation in primary school: A case study. *Journal of Physics: Conference Series*, 948(1). <https://doi.org/10.1088/1742-6596/948/1/012056>
- Retnawati, H., Djidu, H., Kartianom, Apino, E., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st Century*, 76(2), 215–230. <https://doi.org/10.33225/pec/18.76.215>
- Saraswati, P. M. S., & Agustika, G. N. S. (2020). Kemampuan Berpikir Tingkat Tinggi Dalam Menyelesaikan Soal HOTS Mata Pelajaran Matematika. *Jurnal Ilmiah Sekolah Dasar*, 4(2), 257. <https://doi.org/10.23887/jisd.v4i2.25336>
- Sayyah, M., Vakili, Z., Masoudi Alavi, N., Bigdeli, M., Soleymani, A., Assarian, M., & Azarbad, Z. (2012). An Item Analysis of Written Multiple-Choice Questions: Kashan University of Medical Sciences. *Nursing and Midwifery Studies*, 1(2), 83–87. <https://doi.org/10.5812/nms.8738>
- Schraw, G. J., & Robinson, D. H. (2011). *Assessment Of Higer Order Thinking Skillss*. Information Age Publishing.
- Scully, D. (2017). Constructing Multiple-Choice Items to Measure Higher-Order Thinking. *Practical Assessment, Research & Evaluation*, 22(4), 1–13.
- Shete, A., Kausar, A., Lakhkar, K., & Khan, S. (2015). Item analysis: An evaluation of multiple choice questions in Physiology examination. *Journal of Contemporary Medical Education*, 3(3), 106–109. <https://doi.org/10.5455/jcme.20151011041414>
- Shim, G. T. G., Shakawi, A. M. H. A., & Azizan, F. L. (2017). Relationship between Students' Diagnostic Assessment and Achievement in a Pre-University Mathematics Course. *Journal of Education and Learning*, 6(4), 364–371. <https://doi.org/10.5539/jel.v6n4p364>
- Sim, S. M., & Rasiah, R. I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals of the Academy of Medicine Singapore*, 35(2), 67–71.



- Sriwahyuni, K., & Maryati, I. (2022). Kemampuan Pemecahan Masalah Matematis Siswa Pada Materi Program Linear. *Plusminus: Jurnal Pendidikan Matematika*, 2(2), 335–344. <https://doi.org/10.35438/inomatika.v4i1.279>
- Wardhani, S., & Rumiati. (2011). Instrumen Penilaian Hasil Belajar Matematika SMP : Belajar dari PISA dan TIMSS. In *Pusat Pengembangan dan Pemberdayaan Pendidik dan Tenaga Kependidikan (PPPPTK) Matematika*. Pusat Pengembangan dan Pemberdayaan Pendidik dan Tenaga Kependidikan (PPPPTK) Matematika.
- Widana, I. W. (2017). Higher Order Thinking Skills Assessment (HOTS). *JISAE*, 3(1), 32–44.
- Wijaya, A., van den Heuvel-Panhuizen, M., & Doorman, M. (2015). Opportunity-to-learn context-based tasks provided by mathematics textbooks. *Educational Studies in Mathematics*, 89(1), 41–65. <https://doi.org/10.1007/s10649-015-9595-1>
- Wijaya, A., van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *Mathematics Enthusiast*, 11(3), 555–584. <https://doi.org/10.54870/1551-3440.1317>
- Winarso, W. (2014). Membangun Kemampuan Berfikir Matematika Tingkat Tinggi Melalui Pendekatan Induktif, Deduktif Dan Induktif-Deduktif Dalam Pembelajaran Matematika. *Eduma: Mathematics Education Learning and Teaching*, 3(2), 95–118. <https://doi.org/10.24235/eduma.v3i2.58>
- Zohar, A., & Dori, Y. J. (2003). Higher Order Thinking Skills and Low-Achieving Students: Are They Mutually Exclusive? *Journal of the Learning Sciences*, 12(2), 265–272. <https://doi.org/10.1207/S15327809JLS1202>