

Analisis Tingkat Kesukaran Soal Matematika Pada Kemampuan Penyelesaian Masalah Matematis Siswa Kelas VII

Salma Nur Alifah¹, Hamdan Sugilar², Abdurrosyid³

¹Prodi Pendidikan Matematika, UIN Sunan Gunung Djati Bandung
Jl. Cimencrang Kecamatan Gedebage, Kota Bandung, Jawa Barat

^{2,3}Prodi Pendidikan Matematika, Universitas Pendidikan Indonesia
Jl. Dr. Setiabudi No.229, Isola, Kec. Sukasari, Kota Bandung, Jawa Barat

salmaalifhhh@gmail.com

Abstrak

Tujuan penelitian ini adalah untuk menguji validitas, reliabilitas, dan tingkat kesukaran soal matematika yang digunakan untuk menilai kemampuan pemecahan masalah siswa kelas VII salah satu SMP di Kota Bandung. Pengumpulan data dilakukan dengan menggunakan alat ukur yang dibagikan kepada siswa kelas VII. Hasil penelitian menunjukkan bahwa instrumen asesmen memiliki validitas yang sangat rendah sehingga tidak dapat mengukur kemampuan pemecahan masalah siswa secara efektif. Reliabilitas soal esai juga rendah, menandakan soal tersebut perlu diperbaiki.. Sebagian besar soal cukup sulit, namun ada satu soal yang cukup sulit. Selanjutnya, perbedaan kualitas distraktor pilihan ganda terdeteksi. Temuan ini menunjukkan bahwa alat ukur, termasuk pertanyaan yang digunakan, harus direvisi dan diperbaiki untuk memastikan validitas, reliabilitas, dan tingkat kesulitan yang sesuai. Penelitian ini membantu terciptanya alat ukur yang lebih tepat serta soal yang tepat untuk menilai kemampuan siswa dalam memecahkan masalah matematika. Upaya lanjutan untuk menyempurnakan alat ukur akan membantu meningkatkan kualitas pembelajaran matematika di kelas VII dan menghasilkan penilaian keterampilan siswa yang lebih akurat.

Kata kunci: matematika, validitas, reliabilitas

Abstract

The aim of this research was to test the validity, reliability and difficulty level of mathematics questions used to assess the problem solving abilities of class VII students at one of the junior high schools in Bandung City. Data collection was carried out using measuring instruments distributed to class VII students. The research results show that the assessment instrument has very low validity so it cannot measure students' problem solving abilities effectively. The reliability of the essay questions is also low, indicating that the questions need to be improved. Most of the questions are quite difficult, but there is one question that is quite difficult. Furthermore, differences in the quality of the multiple choice distractors were detected. These findings suggest that measurement tools, including the questions used, should be revised and improved to ensure validity, reliability, and appropriate levels of difficulty. This research helps create more precise

measuring tools and appropriate questions to assess students' abilities in solving mathematical problems. Continued efforts to refine measurement tools will help improve the quality of mathematics instruction in seventh grade and result in more accurate assessments of student skills.

Keywords: *mathematics, validity, reliability*

1. PENDAHULUAN

Pendidikan merupakan salah satu cara untuk mengembangkan manusia yang berkompeten dan prospektif. Tingkat pendidikan yang tinggi mencerminkan keberhasilan suatu negara mengadopsi pendidikan formal. Seseorang akan mengembangkan informasi, pemahaman, dan cara berfungsi melalui proses pendidikan (Syah, 2006). Pendidikan yang bermutu tidak dapat dipisahkan dari proses kegiatan pembelajaran di sekolah. Tujuan pengajaran (instructional objective), pengalaman belajar (proses belajar mengajar), dan hasil belajar merupakan tiga komponen proses pembelajaran. Dalam bidang pendidikan, kegiatan penilaian disebut sebagai kegiatan evaluasi. Evaluasi merupakan komponen penting dari proses belajar mengajar. Pendidik mengevaluasi hasil belajar secara berkala untuk mengecek proses, kemajuan, dedikasi, dan perkembangan hasil belajar siswa (Sukardi, 2011). Oleh karena itu evaluasi memegang peranan penting dalam pengembangan sistem pendidikan. Diantara fungsinya adalah memperbaiki sistem dengan masukan dari hasil penilaian, pertanggung jawaban kepada pemerintah, masyarakat atau pihak yang berkepentingan sebagai pihak yang ikut serta mengembangkan pendidikan.

Tujuan diadakannya evaluasi dalam pembelajaran adalah untuk memperoleh informasi yang akurat tentang tingkat godaan tujuan instruksional oleh siswa sehingga dapat mempengaruhi tindakan apa yang akan dilakukan selanjutnya (Aslamiah, 2022). Untuk mencapai tujuan pendidikan tersebut, diperlukan evaluasi tingkat pencapaian tujuan pendidikan. Hal ini sesuai dengan pernyataan House (2012) bahwa tujuan penilaian bukan hanya untuk membuktikan tetapi juga untuk memperbaiki. House melanjutkan, hasil evaluasi akan digunakan untuk mendidik pengambil keputusan tentang apa yang harus dilakukan. Sekarang ada dua organisasi dunia yang menilai kemampuan matematika siswa: Trends in worldwide Mathematics and Science Study (TIMSS) dan Program for International Student Assessment (PISA). Menurut International Association for the Evaluation of Educational Achievement (IEA), Indonesia menempati peringkat ke-45 dari 50 negara peserta survei TIMSS 2015 (IEA, 2015). Sementara itu, menurut temuan penilaian PISA lain yang dikeluarkan oleh Organisasi Kerjasama Ekonomi dan Pembangunan (OECD), Indonesia berada di peringkat ke-69 dari 76 negara peserta pada tahun 2015 (OECD, 2015).

Temuan TIMSS dan PISA menunjukkan bahwa kemampuan matematika siswa Indonesia di tingkat internasional masih jauh di bawah yang diperkirakan. Karena

ujian TIMSS dan PISA terfokus pada masa depan siswa, rendahnya nilai TIMSS dan PISA menunjukkan bahwa siswa Indonesia belum siap menghadapi masalah kehidupan nyata. Penilaian PISA bertujuan untuk mengukur kesiapan generasi muda di akhir sekolah mereka untuk menghadapi tantangan hidup sementara (Anderson, Milford, & Ross, 2008), dan TIMSS mendorong siswa untuk mempertimbangkan pentingnya kesuksesan akademik dan kesuksesan karir di masa depan (Mullis, Martin, Gonzalez, & Chrostowski, 2004). Karena kemampuan menjawab soal-soal matematika merupakan kecakapan hidup yang sangat penting untuk dimiliki siswa, dan kegiatan penilaian merupakan salah satu dari sejumlah dimensi yang sangat penting dalam pendidikan, maka evaluasi kemampuan pemecahan masalah dalam mempelajari matematika sangat diperlukan. Fase kemampuan pemecahan masalah harus diperhatikan saat mengevaluasi bakat ini.

Ada beberapa langkah pemecahan masalah matematika berdasarkan berbagai sumber, namun langkah pemecahan masalah matematika yang paling populer adalah yang dikemukakan oleh Polya, G. (1957), yang meliputi 1) memahami masalah, 2) merencanakan solusi, 3) melaksanakan perhitungan, dan 4) memeriksa kembali proses dan hasil. Masalah validitas adalah apakah butir-butir tes yang dapat dinilai dengan tepat akan diukur. Menurut Taherdoost (2016), validitas dapat didefinisikan sebagai jumlah tes yang mengukur apa yang dirancang untuk diukur. Sullivan (2011) mendefinisikan penentuan validitas suatu alat ukur sebagai "membangun argumen berdasarkan bukti tentang seberapa baik suatu alat ukur mengukur apa yang seharusnya diukur". Sedangkan ketergantungan barang terkait dengan masalah kepercayaan. Reliabilitas adalah suatu tes secara teratur menilai sesuatu yang dapat diandalkan atau dipercaya (Bajpai, 2014). Kaitan antara validitas dan reliabilitas dapat dinyatakan sebagai berikut: validitas sangat penting, tetapi reliabilitas diperlukan karena validitas didukung oleh reliabilitas (Sukiman, 2012). Dalam hal kesulitan soal, Boopathiraj dan Chellamani (2013) mendefinisikannya sebagai fraksi peserta tes yang menjawab soal dengan benar. Tingkat kesulitan soal ditentukan oleh kemampuan atau ketidakmampuan siswa untuk menjawabnya, bukan asumsi pengajar yang membuat soal, karena soal yang sulit atau mudah bagi guru belum tentu sulit atau mudah bagi siswa.

Suatu soal dapat membedakan antara siswa yang mampu (telah menguasai materi yang diujikan) dan siswa yang kurang mampu (belum menguasai materi yang diujikan). Kapasitas hal semacam itu dikenal sebagai daya pembeda. Menurut Kocdar dkk, (2016), daya pembeda suatu soal adalah kemampuan butir soal untuk membedakan antara siswa yang mendapat skor tinggi dan rendah. Dalam hal daya pembeda, soal yang sangat baik adalah soal yang hanya dapat dijawab dengan benar oleh peserta tes yang mampu/pintar/kuasai isi tes dan tidak dapat dijawab dengan benar oleh peserta tes yang belum menguasai topik tes. Peneliti ingin menganalisis validitas, reliabilitas, tingkat kesukaran, dan daya pembeda butir soal yang dapat digunakan untuk mengungkap kemampuan pemecahan masalah siswa dalam penelitian ini. Jumlah alat ukur, dalam contoh ini masalah matematika, mengukur dengan benar apa yang dirancang untuk diukur disebut sebagai validitas.

Validitas yang tinggi berarti alat tersebut secara akurat menilai kemampuan siswa dalam memecahkan masalah matematika. Sejauh mana instrumen menghasilkan data yang konsisten dan dapat diandalkan diukur dengan keandalannya. Kompleksitas pertanyaan mencerminkan kesulitan memecahkan masalah matematika dan dapat digunakan untuk menguji kemampuan siswa. Untuk memenuhi tujuan penelitian, data yang dikumpulkan dari siswa kelas VII di SMPN 11 Bandung akan dianalisis. Informasi ini akan dikumpulkan menggunakan tes matematika yang dibuat khusus dan alat soal terkait. Data tersebut kemudian akan diuji secara statistik untuk mengetahui validitas, reliabilitas, dan tingkat kesukaran soal matematika. Validitas instrumen soal matematika, keterandalannya, dan tingkat kerumitan soal menjadi penekanan utama penelitian ini.

Evaluasi ini harus diteliti secara menyeluruh guna meningkatkan pencapaian tujuan pendidikan di masa mendatang di sekolah. Oleh karena itu, penulis mencoba mengamati sekolah dengan memberikan soal-soal ujian yang kemudian dinilai sebagai salah satu gambaran calon instruktur untuk mengajar nantinya. Sangat penting untuk mempertimbangkan kegunaan tes sebagai alat penilaian. Kualitas tes dapat digunakan untuk menentukan kebenaran hasil belajar siswa. Sebelum menganalisis hasil pembelajaran, pendidik harus melakukan analisis. Analisis adalah kapasitas seseorang untuk menguraikan atau mengkarakterisasi materi atau peristiwa menjadi komponen yang lebih kecil dan memahami hubungan antara satu elemen atau faktor dan variabel lainnya (Sudijono, 2013). Temuan penelitian ini dimaksudkan untuk meningkatkan pengetahuan validitas, reliabilitas, dan tingkat kerumitan soal matematika yang digunakan untuk menilai kemampuan pemecahan masalah siswa kelas VII SMPN 11 Bandung. Selain itu, temuan penelitian ini kemungkinan akan memberikan kontribusi yang signifikan terhadap pengembangan kurikulum dan kemajuan pembelajaran matematika di sekolah.

2. METODE PENELITIAN

Penelitian ini merupakan penelitian deskriptif kuantitatif yang mencoba mendeskripsikan validitas, reliabilitas, tingkat kesukaran, dan daya beda dari butir-butir yang dapat digunakan untuk mengungkapkan kemampuan pemecahan masalah matematis siswa. Untuk menjawab tujuan berikut, item yang dibuat diujicobakan pada 30 siswa dari salah satu SMP di kota Bandung. Item yang dihasilkan diujicobakan kepada 30 siswa untuk menjawab tujuan berikut. Siswa dalam penelitian ini telah mempelajari pola numerik, koordinat Kartesian, hubungan dan fungsi, serta persamaan garis lurus sebelum mengikuti ujian. Sepuluh pertanyaan pilihan ganda diselidiki.

SPSS digunakan untuk menguji validitas soal kemampuan pemecahan masalah matematis. Tingkat korelasi output SPSS dengan tolok ukur berikut untuk memutuskan apakah setiap item valid atau tidak menunjukkan relevansi output SPSS dengan interpretasi kategori validitas item.

Uji reliabilitas butir soal juga dilakukan dengan menggunakan SPSS statistics 22. Nilai reliabilitas butir soal dapat dilihat pada table reliability statistics. Selanjutnya nilai cronbach's alpha dikonsultasikan dengan kriteria reliabilitas yang dikemukakan Hinton (2004) seperti Tabel 1 berikut.

Tabel 1 interpretasi koefisien reabilitas

Besarnya Reliabilitas	Koefisien	Interpretasi
$0,00 < r_{11} < 0,20$		Sangat rendah
$0,20 < r_{11} < 0,40$		Rendah
$0,40 < r_{11} < 0,60$		Cukup
$0,60 < r_{11} < 0,80$		Tinggi
$0,80 < r_{11} < 1,000$		Sangat tinggi

Daya beda merupakan suatu kemampuan butir soal untuk dapat membedakan antara siswa yang berkemampuan tinggi dengan siswa yang (kemampuan rendah). Besarnya daya beda disebut indeks diskriminasi. Semakin tinggi indeks daya beda soal maka semakin mampu soal yang bersangkutan membedakan siswa yang pandai dan siswa yang kurang pandai. Seluruh peserta dikelompokkan menjadi 2 yaitu kelompok atas dan kelompok bawah. Angka yang menunjukkan besarnya daya pembeda adalah disebut Dscriminating Power yang diberi lambang D. Besarnya daya pembeda berkisar antara 0,00 sampai 1,00 (Otaya, 2014)

$$\text{Diskrimination index} = \frac{\text{Average upper group} - \text{Average lower group}}{\text{Full item score}}$$

Tabel 2 interpretasi daya pembed

Index Daya Pembeda	Interpretasi
$D \leq 0,00$	Sangat Jelek
$0,00 < D \leq 0,20$	Jelek
$0,20 < D \leq 0,40$	Cukup
$0,40 < D \leq 0,70$	Baik
$0,70 < D \leq 1,00$	Sangat Baik

Indeks kesukaran atau tingkat kesukaran merupakan bilangan yang menunjukkan sukar dan mudahnya sebuah soal (Johari, 2011). Soal yang baik adalah soal yang

tidak terlalu mudah atau tidak terlalu sukar. Tingkat kesukaran disimbolkan dengan "P" yakni proporsi (Arikunto, 2012).

Tabel 3 Interpretasi indeks kesukaran

Index Kesukaran	Interpretasi
$IK \leq 0,00$	Sangat Sukar
$0,00 < IK \leq 0,20$	Sukar
$0,20 < IK \leq 0,40$	Sedang
$0,40 < IK \leq 0,70$	Mudah
$0,70 < IK \leq 1,00$	Sangat Mudah

Pada soal pilihan ganda terdapat alternative jawaban/option yang merupakan pengecoh (distractor). Butir soal yang baik, pengecohnya akan dipilih secara merata oleh peserta didik yang menjawab salah. Sebaliknya butir soal yang kurang baik, pengecohnya akan dipilih secara tidak merata. Indeks pengecoh dapat dihitung menggunakan rumus (Sugiyono, 2010).

Tabel 4 interpretasi indeks kesukaran

Tingkat Penguasaan	Nilai Akhir	Interpretasi
80% - 100 %	A	Memuaskan
70% - 79%	B	Baik
60% - 69%	C	Cukup
50% - 59%	D	Kurang
0% - 50%	E	Tidak Lulus

3. HASIL DAN PEMBAHASAN

Hasil dan Analisa untuk menguji kemampuan penyelesaian masalah matematis siswa diuji sebagai berikut.

X	Y	XY	X ²	Y ²
---	---	----	----------------	----------------

2518	2482	208257	212898	206318
------	------	--------	--------	--------

Setelah dilakukan perhitungan nilai r_{xy} menggunakan rumus product moment Pearson, diperoleh nilai validitas instrument/bandingannya sebesar 0.10931. Nilai ini menunjukkan bahwa validitas instrumen/bandingannya adalah sangat rendah. Validitas yang rendah mengindikasikan bahwa instrumen yang digunakan dalam menguji kemampuan penyelesaian soal siswa kelas VII di SMPN 11 Bandung tidak secara akurat mengukur apa yang seharusnya diukur.

Validitas instrumen matematika sangat penting untuk memastikan bahwa instrumen tersebut secara akurat mengukur kemampuan siswa dalam memecahkan masalah matematika. Skor validitas yang sangat rendah dalam konteks penelitian ini menunjukkan bahwa instrumen yang digunakan mungkin kurang memadai atau memiliki keterbatasan dalam mengukur kemampuan pemecahan masalah siswa.

Σpq	0,2
Variansi Skor	0,2
KR-20	-0,1
Sangat Rendah	

Setelah melakukan perhitungan reliabilitas untuk soal pilihan ganda, diperoleh nilai reliabilitas sebesar -0,1. Nilai ini menunjukkan bahwa reliabilitas soal PG tersebut sangat rendah. Reliabilitas yang rendah mengindikasikan bahwa soal PG tidak memberikan hasil yang konsisten dan dapat diandalkan dalam mengukur kemampuan penyelesaian soal siswa kelas VII di SMPN 11 Bandung. Dalam penelitian ini, perlu diingat bahwa nilai reliabilitas yang diharapkan harus berada dalam rentang antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan reliabilitas yang lebih baik. Namun, nilai reliabilitas -0,1 menunjukkan adanya masalah serius dalam instrumen soal PG yang digunakan.

Setelah dilakukan perhitungan validitas item pilihan ganda dengan menggunakan r tabel sebesar 0,361, diperoleh hasil bahwa validitas item tersebut tidak memenuhi kriteria yang ditetapkan. Validitas item yang rendah menunjukkan bahwa item pilihan ganda dalam instrumen pengukuran tidak secara akurat mengukur kemampuan penyelesaian soal siswa kelas VII di SMPN 11 Bandung. Validitas item adalah ukuran sejauh mana item-item dalam instrumen pengukuran, dalam hal ini soal pilihan ganda, dapat secara tepat mengukur konstruk yang diinginkan. Nilai validitas item

yang rendah mengindikasikan bahwa item-item tersebut tidak secara konsisten dan akurat menggambarkan kemampuan penyelesaian soal siswa yang sebenarnya.

TK	0,87	1,00	0,97	1,00	1,00	0,00	0,97	1,00	1,00	1,00
interpretasi	Mudah	Mudah	Mudah	Mudah	Mudah	Sangat Sukar	Mudah	Mudah	Mudah	Mudah

Berdasarkan hasil tersebut, dapat disimpulkan bahwa sebagian besar soal dalam instrumen pengukuran ini dianggap mudah oleh siswa. Namun, adanya satu soal yang sangat sulit (TK = 0,00) perlu dipertimbangkan, karena tingkat kesukaran yang terlalu rendah dapat mengurangi kemampuan instrumen dalam mengukur kemampuan penyelesaian soal siswa yang lebih tinggi.

Dalam penelitian selanjutnya, penting untuk memperhatikan keseimbangan antara soal-soal yang mudah dan sulit untuk menghasilkan instrumen yang lebih efektif dalam mengukur kemampuan siswa secara akurat. Evaluasi dan penyesuaian kembali soal-soal yang dianggap sangat mudah atau sangat sulit juga dapat meningkatkan validitas dan reliabilitas instrumen pengukuran.

NO	OPTION			
	A	B	C	D
1	83,3%	6,6%	10%	0%
2	100%	0%	0%	0%
3	0%	96,6%	0%	3,3%
4	0%	0%	100%	0%
5	100%	0%	0%	0%
6	93,3%	0%	23,3%	3,3%
7	96,6%	0%	3,3%	0%
8	100%	0%	0%	0%
9	0%	0%	0%	100%
10	100%	0%	0%	0%

Berdasarkan data di atas, dapat dilakukan analisis kualitas pengecoh pilihan ganda sebagai berikut:

- Soal 1: Opsi A menjadi pilihan yang paling banyak dipilih oleh siswa dengan persentase 83,3%. Sedangkan opsi B, C, dan D memiliki persentase yang jauh

lebih rendah. Hal ini menunjukkan bahwa opsi B, C, dan D kurang efektif sebagai pengecoh karena siswa cenderung memilih opsi A.

- Soal 2, 4, 5, 8, dan 10: Pada kelima soal ini, opsi A menjadi pilihan yang paling banyak dipilih dengan persentase 100%. Ini menunjukkan bahwa opsi B, C, dan D tidak efektif sebagai pengecoh karena siswa tidak memilihnya sama sekali.
- Soal 3, 6, dan 7: Pada ketiga soal ini, terdapat opsi yang memiliki persentase yang cukup tinggi, menunjukkan bahwa pengecohnya cukup efektif. Namun, terdapat pula opsi lain yang memiliki persentase yang signifikan, menunjukkan bahwa masih terdapat siswa yang tergodanya oleh pengecoh yang kurang efektif.
- Soal 9: Opsi D menjadi pilihan yang paling banyak dipilih oleh siswa dengan persentase 100%. Hal ini menunjukkan bahwa opsi A, B, dan C tidak efektif sebagai pengecoh karena siswa tidak memilihnya sama sekali.

Berdasarkan analisis di atas, dapat disimpulkan bahwa terdapat variasi dalam kualitas pengecoh pilihan ganda pada instrumen pengukuran ini. Beberapa opsi pilihan berhasil menjadi pengecoh yang efektif, sementara yang lain kurang efektif karena tidak mampu menarik perhatian siswa atau tidak mampu membingungkan mereka. Untuk meningkatkan kualitas pengecoh, diperlukan perbaikan pada formulasi dan konstruksi opsi pilihan, termasuk penggunaan distraktor yang lebih efektif dan relevan dengan pertanyaan.

4. SIMPULAN

Tujuan penelitian ini adalah untuk menguji validitas, reliabilitas, dan tingkat kesukaran soal matematika yang digunakan untuk menilai kemampuan pemecahan masalah siswa kelas VII. Hasil analisis menunjukkan bahwa alat ukur yang digunakan kurang valid, memiliki nilai validitas instrumen/perbandingan yang sangat rendah. Selanjutnya, reliabilitas soal pilihan memiliki nilai yang tidak diinginkan, hal ini menunjukkan bahwa soal tersebut perlu diperbaiki dan diuji ulang. Sebagian besar soal sangat sederhana, namun ada satu soal dengan tingkat kesulitan yang sangat tinggi yang memerlukan perhatian yang cermat. Ada keragaman dalam keefektifan distraktor dalam distraktor pilihan ganda, dengan beberapa opsi yang dipilih menjadi distraktor yang efektif sementara yang lain kurang berhasil. Berdasarkan temuan tersebut, alat ukur dan soal yang digunakan harus diubah dan disempurnakan untuk menjamin validitas, reliabilitas, dan kesulitan yang memadai. Penelitian ini membantu terciptanya alat ukur yang lebih tepat serta soal yang tepat untuk menilai kemampuan siswa dalam memecahkan masalah matematika. Alat ukur peningkatan akan membantu dalam meningkatkan kualitas pembelajaran matematika kelas VII dan menghasilkan penilaian keterampilan siswa yang lebih akurat.

Selain temuan sebelumnya, beberapa kesimpulan tambahan dapat ditarik dari penelitian ini. Pertama, validitas instrumen/perbandingan yang sangat rendah menunjukkan bahwa alat ukur yang digunakan belum tentu dapat mengukur kemampuan pemecahan masalah matematika siswa kelas VII secara akurat.

Diperlukan upaya untuk memperbaiki instrumen dengan mempertimbangkan validitas isi, validitas konstruk, dan validitas kriteria. Kedua, reliabilitas soal pilihan ganda yang tidak dapat diterima menunjukkan bahwa soal pilihan ganda dalam instrumen pengukuran ini tidak dapat diandalkan secara konsisten untuk mengukur kemampuan siswa. Perlu dilakukan revisi rumusan dan pemberian petunjuk soal, serta penilaian ulang jawaban siswa untuk meningkatkan reliabilitas soal pilihan ganda.

Ketiga, sebagian besar soal cukup mudah, artinya soal yang diberikan belum tentu cukup sulit untuk siswa kelas VII. Disarankan agar soal-soal dengan tingkat kerumitan yang lebih bervariasi, termasuk soal-soal dengan tingkat kesukaran sedang atau berat, dimasukkan untuk memperoleh gambaran yang lebih komprehensif tentang kemampuan pemecahan masalah siswa. Keempat, perbedaan yang signifikan dalam kualitas distraktor pilihan ganda ditemukan. Beberapa kemungkinan bekerja dengan baik sebagai pengalih perhatian, sementara yang lain tidak. Pemeriksaan lebih lanjut terhadap karakteristik siswa dan taktik distraksi yang lebih tepat diperlukan untuk meningkatkan kualitas distraktor. Data tambahan tentang preferensi siswa dan respons mereka terhadap opsi pilihan dapat memberikan wawasan yang lebih dalam untuk meningkatkan kualitas pengecoh pilihan ganda.

Sejumlah kendala mungkin telah mempengaruhi data dan kesimpulan yang dicapai dalam penelitian ini. Salah satunya adalah jumlah sampel siswa kelas VII. Disarankan bahwa sampel yang lebih besar digunakan untuk mencakup kelompok siswa yang lebih besar untuk mendapatkan hasil yang lebih umum. Selain itu, mengumpulkan data tambahan tentang jawaban siswa dan preferensi pilihan pilihan dapat memberikan informasi yang lebih menyeluruh mengenai kualitas distraktor pilihan ganda. Untuk meningkatkan kualitas instrumen pengukuran dan mencapai hasil yang lebih valid dan andal, lebih banyak data harus dikumpulkan dan prosedur penilaian yang lebih efektif harus dikembangkan. Temuan penelitian ini berpotensi meningkatkan kemampuan siswa kelas VII dalam mengukur kemampuan memecahkan masalah matematika dan kualitas pembelajaran matematika.

Referensi

- Anderson, Milford, & Ross. (2008). *Multilevel Modeling with HLM: Taking a Second Look at PISA*. Springer Science Bussines Media.
- Arikunto, S. (2012). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Aslamiah, P. D. (2022). *Pengelolaan Kelas*. Depok: Rajawali Printing.
- Bajpai, S. &. (2014). Goodness of measurement: reliability and validity. *International Journal of Medical Science and Public Health*, 112-116.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International journal of social science & interdisciplinary research*, 189-193.
- Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS explained*, East Sussex. England: Routledge Inc.

- House, E. (2012). *Introduction*. In Kellaghan, T., & Stufflebeam, D. L. (Eds.). *International Handbook of Educational Evaluation: Part One: Perspectives/Part Two: Practice (Vol. 9)*. Springer Science & Business Media.
- IEA. (2015). *International Results in Mathematics*. Boston College: TIMSS & PIRLS International Study Center.
- Kocdar, S., Karadag, N., & Sahin, M. D. (2016). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context. *TOJET: The Turkish Online Journal of Educational*, 15(4).
- Mullis, I. V., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. 140 Commonwealth Avenue, Chestnut Hill, MA 02467: TIMSS & PIRLS International Study Center. Boston College.
- OECD. (2015). *Result in Focus*. PISA: OECD Publishing.
- Polya, G. (1957). *How To Solve It, A New Aspect of Mathematical Method. Second Edition*. . New Jersey: Princeton University Press.
- Sudijono, A. (2013). *Pengantar Evaluasi Pendidikan*. Bandung: Alfa Beta.
- Sugiyono. (2010). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Sukardi. (2011). *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta: Bumi Aksara.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani.
- Sullivan, G. M. (2011). A primer on the validity of assessment instruments . *Journal of Graduate Medical Education*, 119-120.
- Syah, M. (2006). *Psikologi Pendidikan dan Pendekatan Baru*. Bandung: PT Remaja Rosdakaya.
- Taherdoost, H. (2016). *Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research*.