

## INSTRUMENTASI KEMAMPUAN PEMECAHAN MASALAH MATEMATIS: ANALISIS RELIABILITAS, VALIDITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL

Aloisius Loka Son

Universitas Timor, Jln. KM 9 Kel. Sasi, kefamenanu-Timor-NTT, [elson44@ymail.com](mailto:elson44@ymail.com)

Diterima 16 Februari 2019, disetujui 4 Maret 2019, diterbitkan 30 April 2019

Pengutipan: Son, A.L. (2019). Instrumen Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Gema Wiralodra*, Vol 10, No 1, Hal 41 - 52, April 2019

### ABSTRAK

Instrumen penilaian yang baik adalah instrumen yang dapat mengungkap data sesuai dengan kenyataan yang sebenarnya, bersifat tetap, ajeg atau dapat dipercaya, serta memiliki tingkat kesukaran yang bervariasi dan mempunyai indeks diskriminasi yang baik pula. Oleh Karena itu dilakukan penelitian ini untuk menganalisis reliabilitas, validitas, tingkat kesukaran, dan daya beda butir soal, yang digunakan untuk mengungkap kemampuan pemecahan masalah matematis siswa. Soal-soal tes yang diuji merupakan soal dalam bentuk essay sebanyak 4 nomor. Hasil perhitungan menunjukkan bahwa ke-4 soal tes kemampuan pemecahan masalah adalah reliabilitas dalam kategori sedang, serta valid dengan kategori masing-masing adalah tinggi. Indeks kesukaran soal nomor 1 dalam kategori mudah, nomor 2 dan 3 dalam kategori sedang, dan nomor 4 dalam kategori sulit, serta indeks diskriminasi yang cukup baik.

**Kata Kunci:** *Kemampuan Pemecahan Masalah, Instrumen penilaian, Reliabilitas, Validitas, indeks kesukaran, indeks diskriminasi.*

### ABSTRACT

A good assessment instrument is an instrument that can reveal data in accordance with the actual reality, is fixed, steady or reliable, and has varying degrees of difficulty and has a good index of discrimination. Therefore this study was conducted to analyze reliability, validity, level of difficulty, and differentiation of items that can be used to reveal students' mathematics problem solving abilities. The test questions tested are questions in the form of 4 numbers. The calculation results show that the 4 test questions for problem solving ability are reliability in the medium category, and valid with each category is high. The index has difficulty number 1 in the easy category, number 2 and 3 in the medium category, and number 4 in the difficult category, and the discrimination index is quite good.

**Keywords:** *Problem Solving Ability, Instruments of assessment, Reliability, Validity, Difficulty Index, Discrimination Index.*

### PENDAHULUAN

Evaluasi dalam dunia pendidikan sangat penting karena digunakan untuk merumuskan kebijakan yang mendukung terciptanya sumber daya manusia yang kompetitif terhadap era globalisasi saat ini. Itulah sebabnya evaluasi/penilaian merupakan bagian yang tak terpisahkan dalam satu kesatuan proses di setiap lembaga pendidikan. Minimal terdapat dua tujuan diadakan evaluasi yakni untuk mengetahui/membuktikan sejauhmana pemahaman peserta evaluasi setelah berproses, dan hasil evaluasi dijadikan sebagai tolak ukur untuk memperbaiki proses. Hal ini sesuai dengan pernyataan House

(2012) bahwa tujuan dari evaluasi bukan hanya untuk membuktikan tetapi untuk memperbaiki. Lebih lanjut House mengatakan bahwa informasi dari evaluasi akan digunakan untuk memberi tahu para pembuat keputusan tentang apa yang harus dilakukan.

Saat ini terdapat organisasi internasional yang melakukan evaluasi terhadap kemampuan siswa dalam matematika yakni *Trends in International Mathematics and Science Study* (TIMSS) dan *Programme for International Student Assessment* (PISA). *The International Association for the Evaluation of Educational Achievement* (IEA) melaporkan hasil studi TIMSS tahun 2015 bahwa Indonesia berada pada peringkat 45 dari 50 negara peserta (IEA, 2015). Sedangkan hasil studi lainnya dilakukan oleh PISA yang dirilis oleh *Organization for Economic Co-operation and Development* (OECD) menunjukkan bahwa pada tahun 2015, Indonesia berada pada peringkat ke 69 dari 76 negara peserta (OECD, 2015). Laporan hasil TIMSS dan PISA di atas menunjukkan bahwa kemampuan matematika siswa Indonesia pada tingkat Internasional masih jauh dari yang diharapkan. Rendahnya hasil TIMSS dan PISA menggambarkan bahwa siswa Indonesia belum mempunyai kesiapan untuk menghadapi tantangan dalam kehidupan nyata karena penilaian yang dilakukan oleh TIMSS dan PISA berorientasi ke masa depan siswa. penilaian yang dilakukan oleh PISA bertujuan untuk mengukur kesiapan generasi muda diakhir sekolahnya dalam menghadapi tantangan kehidupan temporer (Anderson: 2008), dan melalui TIMSS membuat siswa berpikir akan pentingnya sukses di sekolah dan sukses dalam karier di masa depan (Mullis: 2004).

Tantangan-tantangan dalam kehidupan nyata siswa merupakan masalah yang harus di hadapi. Karenanya, siswa harus dipersiapkan untuk menghadapi proses belajar seumur hidup, belajar memecahkan masalah-masalah nyata, masalah-masalah yang dihadapi siswa di lingkungan masyarakat, dengan tujuan untuk membiasakan dan melatih siswa memecahkan masalah yang dihadapi secara matematis. Sehubungan dengan upaya membiasakan dan melatih siswa memecahkan masalah yang dihadapi, *National Council of Teachers of Mathematics* (NCTM) menegaskan bahwa program pembelajaran harus memungkinkan siswa untuk 1) membangun pengetahuan matematika baru melalui pemecahan masalah, 2) memecahkan masalah yang muncul dalam matematika dan dalam konteks-konteks yang lain, 3) menerapkan dan menyesuaikan berbagai strategi yang sesuai untuk memecahkan masalah, dan 4) memonitor dan merefleksikan proses dari pemecahan masalah matematis (NCTM, 2000).

Apabila siswa dibiasakan dengan proses pembelajaran yang ditawarkan oleh NCTM di atas, maka siswa akan memiliki kemampuan yang baik dalam memecahkan

masalah, baik dalam rutinitasnya di sekolah maupun dalam kehidupan sehari-hari. Kemampuan pemecahan masalah adalah kemampuan yang berkaitan dengan proses mengorganisasikan konsep dan keterampilan ke dalam pola aplikasi baru untuk mencapai suatu tujuan. Zevenbergen (2004) menyatakan bahwa dalam memecahkan masalah, siswa perlu memiliki pemahaman dan pengetahuan yang memadai, serta memiliki berbagai macam strategi yang dapat dipilih ketika menghadapi masalah yang berbeda. Pemecahan masalah merupakan salah satu elemen penting yang menggabungkan masalah kehidupan nyata dan pengalaman. Oleh karena itu, kemampuan pemecahan masalah sangat perlu untuk dikembangkan serta dikuasai oleh siswa, seperti yang dikemukakan oleh Anderson (2009) bahwa pemecahan masalah merupakan kecakapan hidup yang penting untuk dikuasai siswa.

Karena kemampuan pemecahan masalah matematis adalah kecakapan hidup yang penting dikuasai siswa, dan kegiatan evaluasi merupakan salah satu dimensi dari sejumlah dimensi yang sangat penting dalam pendidikan, maka perlu diadakan evaluasi terhadap kemampuan pemecahan masalah dalam pembelajaran matematika. Evaluasi terhadap kemampuan ini harus memperhatikan langkah-langkah kemampuan pemecahan masalah. Terdapat sejumlah langkah pemecahan masalah matematis berdasarkan sumber yang berbeda-beda, namun langkah-langkah pemecahan masalah matematis yang populer adalah langkah-langkah yang dikemukakan oleh Polya, G. (1957) antara lain 1) memahami masalah, 2) merencanakan penyelesaian, 3) melaksanakan perhitungan, dan 4) memeriksa kembali proses dan hasil.

Kegiatan mengukur kemampuan pemecahan masalah matematis dapat diibaratkan sebagai kegiatan memotret. Dalam memotret memerlukan alat potret yang disebut dengan instrument penilaian, sedangkan gambar hasil potretan dalam kegiatan penilaian dikenal dengan data penilaian. Data yang baik adalah data yang sesuai dengan kenyataan yang sebenarnya, dan data tersebut bersifat tetap, ajeg atau dapat dipercaya. Oleh karena itu, instrumen penilaian harus memiliki bukti validitas, reliabilitas, tingkat kesukaran dan daya pembeda butir soal. Keempat bukti di atas merupakan analisis butir soal secara kuantitatif.

Validitas berkaitan dengan permasalahan apakah soal tes yang dimaksudkan dapat mengukur secara tepat sesuatu yang akan diukur tersebut. Taherdoost (2016) mengatakan bahwa validitas dapat diartikan sebagai sejauh mana tes itu mengukur apa yang seharusnya diukur. Menentukan validitas alat ukur dapat dipandang sebagai membangun argument berbasis bukti tentang seberapa baiknya alat ukur mengukur apa yang seharusnya diukur (Sullivan, 2011). Sedangkan reliabilitas butir soal berhubungan dengan masalah

kepercayaan. Reliabilitas menggambarkan bahwa suatu tes mengukur sesuatu secara konsisten yang dapat diandalkan atau dipercaya (Bajpai, 2014). Hubungan antara validitas dan reliabilitas dapat dijelaskan bahwa validitas itu penting, sedangkan reliabilitas itu perlu, karena reliabilitas itu menyokong validitas (Sukiman, 2012). Sehubungan dengan tingkat kesukaran butir soal, Boopathiraj & Chellamani (2013) mendefinisikannya sebagai proporsi peserta tes yang menjawab soal tersebut dengan benar. Tingkat kesukaran butir soal dipandang dari kesanggupan atau kemampuan siswa dalam menjawabnya, bukan dari asumsi guru yang menyusun soal, karena butir soal yang sulit atau mudah bagi guru belum tentu sulit atau mudah bagi siswa. Suatu butir soal dapat membedakan antara siswa yang mampu (menguasai materi yang ditanyakan) dan siswa yang kurang mampu (belum menguasai materi yang ditanyakan). Kemampuan suatu butir soal yang demikian disebut daya beda (diskriminasi). Kocdar, et al (2016) mendefinisikan bahwa daya beda (diskriminasi) suatu soal merupakan kemampuan item soal untuk membedakan siswa yang mendapat skor tinggi dan skor rendah. Dalam kaitannya dengan daya pembeda, soal yang baik adalah soal yang dijawab benar oleh peserta tes yang mampu/pandai/menguasai materi tes, dan tidak dapat dijawab secara benar oleh peserta tes yang belum menguasai materi tes.

Analisis terhadap validitas, reliabilitas, tingkat kesukaran, dan daya beda dimaksudkan untuk mengungkap kualitas butir soal, sehingga dalam penelitian ini, peneliti ingin menganalisa validitas, reliabilitas, tingkat kesukaran, dan daya beda butir soal yang dapat digunakan untuk mengungkap kemampuan pemecahan masalah siswa.

## **METODOLOGI PENELITIAN**

Penelitian ini merupakan jenis penelitian deskriptif yang bertujuan untuk mendeskripsikan validitas, reliabilitas, tingkat kesukaran, dan daya beda butir soal yang dapat digunakan untuk mengungkap kemampuan pemecahan masalah matematis siswa. Untuk menjawab tujuan di atas, butir soal yang telah didesain diujikan terhadap 19 siswa SMPN I Kefamenanu pada bulan oktober 2018. Siswa yang mengikuti tes dalam penelitian ini adalah mereka yang telah belajar tentang materi pelajaran operasi bentuk aljabar. Soal-soal yang diujikan adalah soal dalam bentuk essay yang berjumlah 4 nomor.

Uji validitas butir soal kemampuan pemecahan masalah matematis ini dilakukan dengan menggunakan SPSS statistics 22. Valid dan tidaknya setiap butir soal ditentukan dengan membandingkan signifikansi pada output SPSS dengan  $\alpha = 0,05$ . Sedangkan

interpretasi kategori validitas butir soal dapat dikonsultasikan besarnya korelasi pada output SPSS dengan patokan berikut.

**Tabel 1.**  
**Kriteria koefisien korelasi *product moment***

No	$r_{xy}$	Kategori
1	$0,80 \leq r_{xy} \leq 1,00$	Sangat tinggi
2	$0,60 \leq r_{xy} < 0,80$	Tinggi
3	$0,40 \leq r_{xy} < 0,60$	Cukup
4	$0,20 \leq r_{xy} < 0,40$	Rendah
5	$0,00 \leq r_{xy} < 0,20$	Sangat rendah

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 2 berikut.

**Tabel 2.**  
**Kriteria Reliabilitas butir soal**

No	Koefisien Reliabilitas	Penafsiran
1	$0,00 \leq r_i < 0,50$	Derajat reliabilitas rendah
3	$0,50 \leq r_i < 0,70$	Derajat reliabilitas sedang
3	$0,70 \leq r_i < 0,90$	Derajat reliabilitas tinggi
4	$0,90 \leq r_i \leq 1,00$	Derajat reliabilitas sangat tinggi

Uji tingkat kesukaran butir soal kemampuan matematis ini menggunakan rumus:

$$Difficulty\ index = \frac{Average\ score}{Full\ item\ score} \dots\dots\dots (Johari, 2011)$$

Kriteria indeks kesukaran butir soal tes kemampuan pemecahan masalah matematis siswa dapat dikonsultasikan dengan Tabel 3. berikut.

**Tabel 3.**  
**Kriteria Indeks Kesukaran Butir soal**

No	Indeks Kesukaran	Kriteria
1	$0,00 \leq DI < 0,30$	Soal sukar
2	$0,30 \leq DI < 0,80$	Soal sedang

3	$0,80 \leq DI \leq 1,00$	Soal mudah
---	--------------------------	------------

Daya beda soal adalah kemampuan suatu soal untuk membedakan antara siswa yang berkemampuan tinggi dengan siswa yang berkemampuan rendah. Angka yang menunjukkan besarnya daya beda disebut indeks diskriminasi (D). Indeks diskriminasi butir soal uraian dapat dihitung dengan rumus perbandingan antara selisih mean kelompok atas dan mean kelompok bawah dengan skor maksimal tiap butir soal (Boopathiraj & Chellamani, 2013) yaitu:

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

Selanjutnya dikonsultasikan harga  $D$  dengan kriteria seperti pada Tabel 4. berikut.

**Tabel 4.**  
**Kriteria Indeks Daya Pembeda Butir Soal**

No	IDP	Interpretasi
1	Tanda negative	Tidak ada daya pembeda
2	$0,00 \leq D < 0,20$	Lemah
3	$0,20 \leq D < 0,40$	Cukup
4	$0,40 \leq D < 0,70$	Baik
5	$0,70 \leq D \leq 1,00$	Bai ksekali

## HASIL PENELITIAN DAN PEMBAHASAN

### Hasil Penelitian

Hasil analisa butir soal tes kemampuan pemecahan masalah matematis seperti uraian berikut ini.

#### 1. Validitas Buti Soal

Validitas butir soal tes kemampuan pemecahan masalah matematis diuji menggunakan SPSS statistics 22, dengan outputnya seperti pada table 5 berikut.

**Table 5.**  
**Validitas Butir Soal Tes Kemampuan Pemecahan Masalah**

No	Correlation					Ket.	Kategori
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>tot</sub>		
1	X <sub>1</sub>	Pearson Correlation Sig. (2-tailed)	1			0.729** 0.000	Valid Tinggi
2	X <sub>2</sub>	Pearson Correlation Sig. (2-tailed)		1		0.746** 0.000	Valid Tinggi
3	X <sub>3</sub>	Pearson Correlation Sig. (2-tailed)			1	0.654** 0.002	Valid Tinggi
4	X <sub>4</sub>	Pearson Correlation Sig. (2-tailed)				0.792** 0.000	Valid Tinggi
	N		19			19	

Tabel 5 di atas menunjukkan bahwa butir soal nomor 1-4 yang disimbolkan dengan  $X_1$ - $X_4$  adalah valid dengan kategori validitasnya masing-masing adalah tinggi.

## 2. Reliabilitas Butir Soal

Reliabilitas butir soal tes kemampuan pemecahan masalah diuji menggunakan SPSS statistics 22, yang outputnya seperti pada table 6 berikut.

**Table 6.**

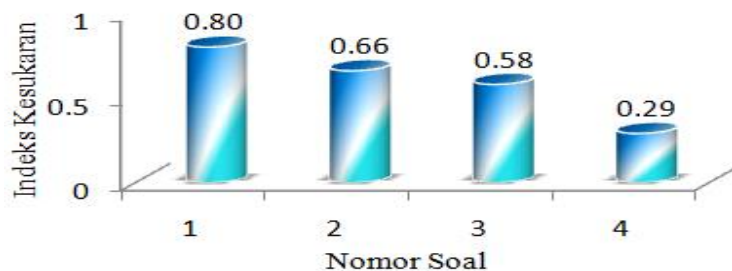
### Reliabilitas Butir Soal Tes Kemampuan Pemecahan Masalah (TKPM)

Reliabilitas Statistics				
Kemampuan Matematis	Cronbach's Alpha	N of Item	Keterangan	Kriteria
TKPM	0.699	4	Reliabel	Sedang

Nilai cronbach's alpha pada table 6 sebesar 0.699 sehingga disimpulkan bahwa soal tes kemampuan pemecahan masalah matematis adalah reliabel, dengan kategori sedang.

## 3. Tingkat Kesukaran Butir Soal

Hasil perhitungan indeks tingkat kesukaran butir soal Tes Kemampuan Pemecahan Masalah (TKPM) seperti terlihat pada diagram di berikut ini.



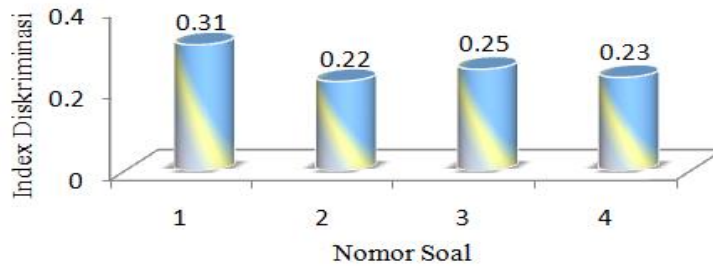
Gambar 1.

### Indeks Kesukaran Butir Soal

Diagram 1 menunjukkan bahwa soal TKPM nomor 1 dalam kategori mudah, nomor 2 dan 3 dalam kategori sedang, dan nomor 4 dalam kategori sulit. Hasil ini menunjukkan bahwa proporsionalitas dari keempat butir soal TKPM adalah normal yakni 25% dalam kategori mudah, 50% dalam kategori sedang, dan 25% dalam kategori sukar.

## 4. Daya Pembeda Butir Soal

Indeks daya pembeda butir soal Tes Kemampuan Pemecahan Masalah seperti terlihat pada diagram di berikut ini.



Gambar 2.

#### Indeks Diskriminasi Butir Soal

Diagram di atas menunjukkan bahwa index diskriminasi keempat soal TKPM berada pada interval  $0,20 \leq D < 0,40$  yakni dalam kategori cukup baik (satisfactory).

#### *Pembahasan*

Bentuk soal yang dianalisa dalam penelitian ini berupa soal-soal pemecahan masalah dalam bentuk uraian dengan pedoman pemecahannya disesuaikan dengan tahapan pemecahan masalah yang diusulkan Polya (1957). Alasan menggunakan langkah pemecahan masalah tersebut adalah langkah-langkah tersebut simpel dan sederhana sehingga mudah dalam menganalisis jawaban siswa.

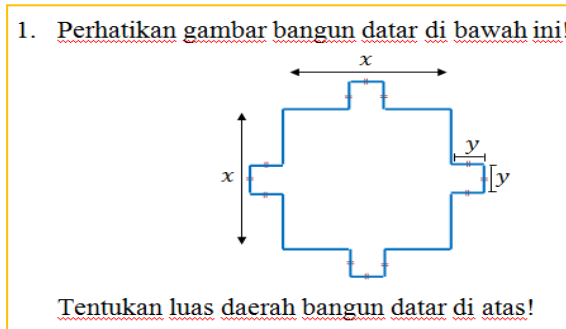
Butir soal tes yang dianalisa, mengukur kemampuan siswa SMP dalam memecahkan masalah yang berkaitan dengan materi pelajaran Operasi Bentuk Aljabar, yang cakupan luas materinya masih sebatas memahami bentuk aljabar serta operasi penjumlahan, pengurangan, perkalian, dan pembagian bentuk aljabar. Keempat soal yang didesain secara keseluruhan memiliki kualitas yang baik dari segi reliabilitas, validitas, indeks kesukaran, dan indeks daya beda.

Hasil perhitungan reliabilitasnya memberikan nilai cronbach's alpha sebesar 0.699. Besarnya nilai cronbach's alpha seperti di atas dikategorikan dalam butir soal yang reliabel dengan kategori sedang, karena 0.699 berada pada interval 0.50-0.70 (Hinton, 2004). Hasil ini menunjukkan bahwa butir soal tes kemampuan pemecahan masalah dapat dipercaya karena cenderung memberikan hasil yang tetap. Walaupun telah dinyatakan reliable namun belum cukup dan harus dikombinasikan dengan validitas, karena soal tes yang reliable belum tentu valid (Sukiman, 2012).

Analisis reliabilitas dilakukan terhadap sejumlah butir soal tes secara keseluruhan, namun analisis validitas dilakukan pada masing-masing butir soal tes. Validitas butir soal adalah ketepatan mengukur yang dimiliki oleh sebutir soal yang merupakan bagian tak terpisahkan dari tes sebagai suatu totalitas (Sukiman, 2012). Sama halnya dengan tingkat



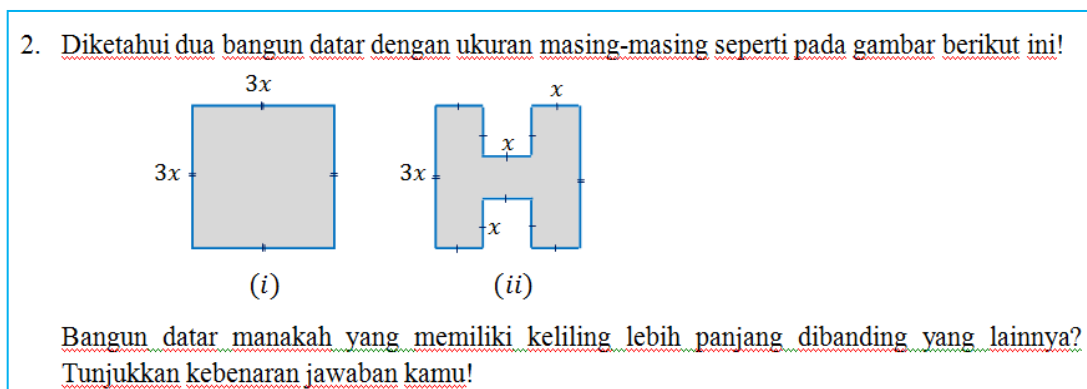
kesukaran serta daya beda butir soal yang perhitungannya secara sendiri-sendiri. Uji validitas, indeks kesukaran, dan daya beda dari keempat soal seperti uraian berikut.



Gambar 3.  
Petikan Butir Soal Nomor 1.

Output SPSS dalam perhitungan validitas butir soal nomor 1 di samping, memberikan nilai Sig. (2-tailed) sebesar  $0.00 < 0.05$  yang berarti butir soal tersebut adalah valid. Sedangkan nilai pearson correlation sebesar  $0.729^{**}$  yang menunjukkan bahwa kategori validitas butir soal nomor 1 di samping adalah tinggi, karena berada pada interval 0.60-0.80 (Sukiman, 2012). Hasil difficulty indexnya sebesar 0.80 (mudah) serta discrimination indexnya sebesar 0.31 dalam kategori cukup baik.

Butir soal nomor 2 mengukur kemampuan siswa dalam menjumlahkan bentuk aljabar sehubungan dengan mencari keliling bangun datar seperti petikan berikut.



Gambar 4.  
Petikan Butir Soal Nomor 2.

Output SPSS dalam perhitungan validitas butir soal nomor 2 di atas memberikan nilai Sig. (2-tailed) sebesar  $0.00 < 0.05$ . Secara statistik, hal ini berarti butir soal nomor 2 adalah valid. Nilai pearson correlation sebesar  $0.746^{**}$  berada pada interval 0.60-0.80 sehingga validitas butir soal nomor 2 dalam kategori tinggi. Sedangkan hasil *difficulty index* sebesar

0.66 (sedang) serta *discrimination index* sebesar 0.22 dalam kategori cukup baik karena berada pada interval  $20 \leq D < 40$  (Boopathiraj & Chellamani, 2013).

Indikator soal nomor 3 mengukur kemampuan siswa dalam merumuskan serta menjumlahkan bentuk aljabar seperti petikan berikut.

3. Seorang guru matematika merahasiakan tiga bilangan kepada murid-muridnya. Dia hanya memberi tahu jumlah dari masing-masing dua bilangan tersebut adalah 20, 32, dan 40. Berapakah jumlah ketiga bilangan tersebut?

Gambar 5.  
Petikan Butir Soal Nomor 3.

Nilai Sig. (2-tailed) pada output SPSS dalam perhitungan validitas butir soal nomor 3 sebesar  $0.00 < 0.05$ , serta nilai *pearson correlation* sebesar  $0.654^{**}$ . Nilai statistik ini menunjukkan bahwa butir soal nomor 3 adalah valid dengan kategori tinggi. Hasil perhitungan *difficulty index* sebesar 0.58 dalam kategori sedang, karena berada pada interval  $0.30 \leq DI < 0.80$  (Johari, 2011) serta *discrimination index* sebesar 0.25 dalam kategori cukup baik.

Butir soal nomor 4 mengukur kemampuan siswa dalam menghitung hasil bagi bentuk aljabar sehubungan dengan mencari luas persegi panjang seperti petikan berikut.

4. Suatu kebun berbentuk persegi panjang dengan sisi panjangnya adalah  $(4x + 2)$  m dan luasnya  $(12x^2 + 2x - 2)$  m<sup>2</sup>.  
a. Tentukan lebar kebun tersebut!  
b. Apakah nilai  $x = 7$  m mengakibatkan lebar kebun tersebut adalah 20 m, panjang 30 m, dan luas 600 m<sup>2</sup>? Tunjukkan kebenaran jawaban kamu!

Gambar 6.  
Petikan Butir Soal

Nilai Sig. (2-tailed) pada output SPSS dalam perhitungan validitas butir soal nomor 4 sebesar  $0.00 < 0.05$ , serta nilai *pearson correlation* sebesar  $0.792^{**}$ . Secara statistik, hal ini berarti butir soal nomor 4 adalah valid, dengan kategori validitas tinggi. Sedangkan hasil perhitungan *difficulty index* sebesar 0.29 dalam kategori sulit, dan hasil *discrimination index* sebesar 0.23 dalam kategori cukup baik.

Berdasarkan uraian hasil penelitian dan pembahasan di atas dapat disimpulkan bahwa keempat butir soal tes tersebut mempunyai kualitas yang baik dari segi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda serta layak digunakan untuk mengukur kemampuan pemecahan masalah matematis siswa.

## PENUTUP

Berdasarkan hasil analisis dan pembahasan di atas, dapat disimpulkan bahwa:

- 1) Keempat butir soal tes kemampuan pemecahan masalah matematis siswa adalah valid dengan kategori validitas masing-masing adalah tinggi.
- 2) Soal tes kemampuan pemecahan masalah matematis adalah reliabel, dengan kategori sedang, yang dapat dilihat pada nilai cronbach's alpha dalam perhitungan reliabilitas sebesar 0.699.
- 3) Indek kesukaran soal nomor 1 sebesar 0.80 dalam kategori mudah, nomor 2 sebesar 0.66 dalam kategori sedang, nomor 3 sebesar 0.58 dalam kategori sedang, dan nomor 4 sebesar 0.29 dalam kategori sulit.
- 4) Index diskriminasi soal nomor 1 sebesar 0.31, soal nomor 2 sebesar 0.22, soal nomor 3 sebesar 0.25, dan soal nomor 4 sebesar 0.23. Keempat soal TKPM ini memiliki memiliki daya beda cukup baik (satisfactory).

## DAFTAR PUSTAKA

- Anderson, J. (2009). *Mathematics Curriculum Development and the Role of Problem Solving*. ACSA Conference 2009.
- Anderson, Milford, and Ross (2008). Multilevel Modeling with HLM: Taking a Second Look at PISA. In Shelley, M. C., Yore, L. D., & Hand, B. B. (Eds.). *Quality research in literacy and science education: international perspectives and gold standards*. Springer Science & Business Media.
- Bajpai, S., & Bajpai, R. (2014). Goodness of measurement: reliability and validity. *International Journal of Medical Science and Public Health*, 3(2), 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*, 2(2), 189-193.
- Hinton, P. R., Brownlow, C., McMurray, I. & Cozens, B. (2004). *SPSS explained*, East Sussex, England, Routledge Inc.
- House, E.R. (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*. TIMSS & PIRLS International Study Center: Lynch School of Education, Boston College.

- Johari, J., Sahari, J., Wahab, D. A., Abdullah, S., Abdullah, S., Omar, M. Z., & Muhamad, N. (2011). Difficulty index of examinations and their relation to the achievement of programme outcomes. *Procedia-Social and Behavioral Sciences*, 18, 71-80.
- 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 Technology*, 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*. TIMSS & PIRLS International Study Center. Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467.
- NCTM. (2000). *NCTM Program standards 2000, Programs for Initial preparation of Mathematics Teachers*. Reston. VA: NCTM.
- OECD. (2015). *Result in Focus. PISA*, OECD Publishing.
- Polya, G. (1957), *How To Solve It, A New Aspect of Mathematical Method*. Second Edition. Princeton University Press: New Jersey.
- Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*. 3:119-120.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi*. Insan Madani: Yogyakarta.
- Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research.
- Zevenbergen, R., Dole, S., and Wright, R.J. (2004). *Teaching Mathematics in Primary School*. Sidney: Allen and Unwin.