



## Testing geometry problem items with PCM scales for secondary school students in Samarinda

### Pengujian butir soal geometri dalam penskalaan PCM tingkat SMP di Samarinda

Sugeng<sup>1\*</sup>, Kukuh<sup>2</sup>, Jumriah<sup>3</sup>

<sup>1,2,3</sup> Pendidikan Matematika FKIP Universitas Mulawarman, Samarinda, Indonesia

\* Email Penulis Korespondensi: [sugeng.medina01@gmail.com](mailto:sugeng.medina01@gmail.com)

| Article Information  | Abstract   |
|--|--|
| <b>Keywords:</b><br>Geometry problem<br>PCM Scales<br>Testing      | <i>This study aims to reveal the characteristics of the geometry problem in scaling the Partial Credit Model (PCM) at the junior high school level in Samarinda in 2022. This study involved students of SMPN 2 Samarinda and SMPN 8 Samarinda (N=122) as samples. Testing of 40 problems of Geometry using the Quest application. The preparation of Geometry questions goes through stages ranging from compiling test specifications, writing test questions, and studying test questions. The test results show that the overall question items are fit and valid; its reliability coefficient is indicated by Internal Consistency 70; all persons are also fit for the model, but there are 19 persons to consider. The use of this instrument in measurement will be more accurate if further relevant research considers the broader scope of the person.</i>  |
| Info Artikel   | Abstrak  |
| <b>Kata kunci:</b><br>Soal Geometri<br>Penskalaan PCM<br>Pengujian | Penelitian ini bertujuan untuk mengungkap karakteristik butir soal Geometri dalam penskalaan <i>Partial Credit Model (PCM)</i> pada jenjang SMPN di Samarinda tahun 2022. Penelitian ini melibatkan siswa SMPN 2 Samarinda dan SMPN 8 Samarinda (N=122) sebagai sampel. Pengujian terhadap 40 butir soal Geometri menggunakan aplikasi Quest. Penyusunan soal Geometri (dalam penskalaan <i>PCM</i> ) melalui tahapan mulai dari menyusun spesifikasi tes, menulis soal tes, dan menelaah soal tes. Hasil pengujian menunjukkan bahwa keseluruhan butir soal adalah <i>fit</i> dan valid; koefisien reliabilitasnya ditunjukkan dengan <i>Internal Consistency 70</i> ; seluruh person juga <i>fit</i> terhadap model, namun terdapat 19 person yang harus dipertimbangkan. Penggunaan Instrumen ini pada pengukuran akan menjadi lebih akurat, apabila dalam penelitian lebih lanjut yang relevan mempertimbangkan cakupan personnya lebih luas |

Copyright (c) 2022 The Author  
This is an open access article under the CC-BY-SA license



## PENDAHULUAN

Salah satu penentuan kualitas hasil penelitian adalah kualitas data. Data yang berkualitas memberikan hasil analisis yang dapat dipecahkan. Untuk memperoleh data yang berkualitas, peneliti memerlukan adanya instrumen yang juga berkualitas. Kondisi demikian, menunjukkan bahwa instrumen yang berkualitas sangat menentukan kualitas data dan keakuratan hasil pengukuran (penelitian). Djemari Mardapi (2012), menyatakan bahwa “tingkat keberhasilan suatu program dapat diketahui melalui suatu pengukuran; termasuk prestasi belajar siswa dapat diketahui melalui kegiatan pengukuran”. Kondisi demikian menunjukkan bahwa untuk dapat melaksanakan kegiatan pengukuran maka perlu ketersediaan tes prestasi belajar atau alat ukur (instrumen) sejenis yang berkualitas, terutama pada bidang pembelajaran matematika.

Menurut beberapa ahli bidang pengukuran pendidikan, diungkapkan bahwa pengukuran merupakan (a) penetapan angka dengan cara yg sistematis untuk menyatakan keadaan suatu objek/individu (Allen & Yen, 1979); (b) sebagai sekumpulan aturan untuk menetapkan suatu bilangan yg mewakili objek, sifat, atau karakteristik, atribut atau tingkah laku (Reynold, et al. 2010, Kusaeri & Suprananto, 2012). Dalam pengukuran memerlukan adanya fenomena atau gejala yang menjadi atribut, karakteristik dari suatu objek/individu untuk kemudian ditetapkan suatu angka dengan aturan yang sistematis.

Pendekatan dalam pengukuran terdapat dua macam, yaitu pendekatan teori klasik (*CTT; Classical Theory Test*) dan pendekatan teori tes modern (*IRT; Item Response Theory*) (Hambleton, Swaminathan, & Rogers, 1991). *CTT* sebagai pendekatan pengukuran yang keberadaannya lebih awal daripada pendekatan yang kedua. Teori ini bekerja berdasarkan kelompok individu. Menurut Sumadi Suryabrata (2000); Djemari Mardapi (2012), Saifuddin Azwar (1999, 2012); bahwa terdapat beberapa asumsi dalam teori tes klasik ataupun teori tes modern. Pada teori tes klasik antara lain, (1) skor perolehan ( $X$ ) terdiri dari skor murni ( $X_f$ ) dan skor kesalahan pengukuran ( $X_e$ ); (2) nilai harapan skor perolehan samadengan skor murni; (3) skor murni dan skor kesalahan yang dicapai oleh suatu populasi subjek pada suatu tes tidak berkorelasi satu sama lain, dan lainnya. Dalam teori tes modern, terdapat dua asumsi utama, yaitu (a) asumsi unidimensi (*unidimensionality*), maksudnya untuk model-model *IRT* berlaku hanya ada satu kemampuan yang diukur oleh butir-butir soal dalam suatu perangkat tes; atau dikatakan substansi yang diukur adalah satu dimensi. Hambleton, Swaminathan, & Rogers (1991) mengungkapkan bahwa asumsi unidimensi dalam pengukuran dimaksudkan hanya satu kemampuan yang diukur dengan butir-butir tes. Asumsi kedua adalah (b) independen lokal (ketidak-tergantungan lokal; *local independence*), maksudnya probabilitas menjawab benar butir soal satu dengan lainnya adalah independen.

Dalam penelitian ini, peneliti tertarik untuk melakukan suatu pengadaan instrumen tes bidang Geometri menurut teori tes modern. Salah satu model penskoran atau penskalaan dalam teori tes modern adalah *Partial Credit Model (PCM)*. Sebagai salah satu model penskalaan dalam *IRT*, *PCM* merupakan pengembangan dari model Rasch (alternatif jawaban yang disediakan bersifat dikotomi) menjadi model dengan option yang lebih dari dua (politomi) dan dikembangkan oleh Masters (1982). sebagai model pengukuran untuk menskala soal-soal bentuk respons terkonstruksi (Fitzpatrick, et al., 1996); untuk penskalaan pada perangkat tes (Sykes & Yen, 2000); yang memiliki ketelitian dan efisiensi dalam praktik pengukuran (Baker, Rounds, & Zevon, 2000). *Partial Credit Model* memberikan kredit (penghargaan, apresiasi, kredit) kepada testi yang menjawab soal. Bentuk soal *PCM* mirip dengan soal *multiple choice*, namun menyediakan alternatif jawaban lebih dari dua. Model ini diterapkan dalam pengadaan soal geometri dengan tujuan memberikan pengenalan tentang bentuk soal yang relatif baru bagi guru-guru agar memiliki variasi dalam pengadaan soal hasil belajar. Selain itu, juga menjadi tantangan bagi guru-guru untuk berinovasi dalam pengukuran hasil belajar.

Geometri merupakan salah satu cabang Matematika yang mempelajari bentuk

bangun- bangun datar dan ruang, beserta ukuran-ukuran atas bagian-bagiannya (titik, garis, bidang, sudut, dan lainnya). Korespondensi antara titik-titik tertentu pada bidang menimbulkan adanya transformasi. Kondisi demikian, pada saat membahas soal koordinat titik pada dimensi dua, akan berkenaan dengan proses transformasi (Eccles, terjemahan Sudrajat, 2003). Materi geometri yang dipelajari siswa SMP sebagaimana termuat dalam Kurikulum 2013.

## METODE

Penelitian jenis deskriptif ini menggunakan subjek penelitian sebanyak empat kelas siswa SMP Negeri Samarinda (SMPN 2 dan SMPN 8 Samarinda). Keempat kelas tersebut berada pada dua sekolah berbeda yang setiap sekolah masing-masing terdiri dari dua kelas. Kedua sekolah ini dianggap mewakili sekolah berdasarkan letak geografisnya (sampel bersifat purposif), yaitu tengah kota dan pinggiran kota Samarinda. Total subjek ada 122 siswa. Instrumen penelitian berupa butir soal tes Geometri dalam penskalaan *PCM*, yang disusun oleh peneliti berdasarkan kurikulum 2013. Dalam pengadaan soal atau instrumen Geometri pada penelitian ini mencakup materi pada kelas VIII, yang terdiri atas: Koordinat Cartesius, Teorema Pythagoras, Lingkaran, Bangun Ruang Sisi Datar, Segiempat dan Segitiga, Kesebangunan dan Kekongruenan, dan Bangun Ruang Sisi Lengkung.

Sebelum dilakukan pengujian instrumen, soal yang dimaksud harus diproduksi dulu. Langkah penyusunan soal, antara lain (1) menyusun spesifikasi tes, (2) menulis tes, (3) menelaah tes, (4) melakukan ujicoba tes, dan (5) menganalisis butir soal (Djemari Mardapi, 2008, 2012; Sumadi Suryabrata, 2000). Dalam penyusunan spesifikasi tes, mencakup beberapa kegiatan, yaitu (a) rekaan teoretis, (b) menentukan subjek, (c) tujuan pengukuran, (d) materi tes, (e) tipe soal, (f) jumlah soal, dan (g) kisi-kisi soal (Sumadi Suryabrata, 2000).

Dalam penyusunan soal cenderung dilakukan oleh orang-orang yang kompeten pada bidangnya (misal guru). Untuk menyusun kisi-kisi tes ditinjau dari arah isi, format, dan pembahasan. Pelaksanaan uji coba (pengujian) soal dilakukan pada subjek yang relevan, dengan mempertimbangkan alokasi waktu yang tersedia, dan memilih subjek yang tidak hanya berada pada satu peringkat tertentu. Selain alokasi waktu, juga mempertimbangkan spesifikasi petunjuk pengerjaan soal, kondisi pengujian, dan kedisiplinan dalam mengerjakan soal.

Analisis butir soal tes meliputi penentuan karakteristik soal, agar diperoleh butir soal yang berkualitas secara empirik. Untuk mampu menganalisis butir soal dengan teori tes modern (*IRT*), perlu memahami terlebih dahulu karakteristik pada teori tes klasik, sebagai dasar pengetahuan pengukuran. Menurut teori tes modern, analisis butir soal *PCM*, dapat dilakukan menggunakan aplikasi, antara lain dengan Quest (Adams & Khoo, 1996). Karakteristik soal, antara lain mencakup *difficulty level* (tingkat kesukaran), kecocokan butir soal terhadap model pengukuran (*items fit*), kecocokan person terhadap model pengukuran (*person fit*). Butir *PCM*, sebagai bentuk pengembangan model Rasch, dan hanya memiliki parameter tingkat kesukaran dan untuk karakter daya pembeda dianggap sama, yaitu 1 (Embretson & Reise, 2010). Oleh karenanya, karakter daya pembeda pada butir *PCM* tidak menjadi pertimbangan untuk dianalisis lebih lanjut.

## HASIL DAN DISKUSI

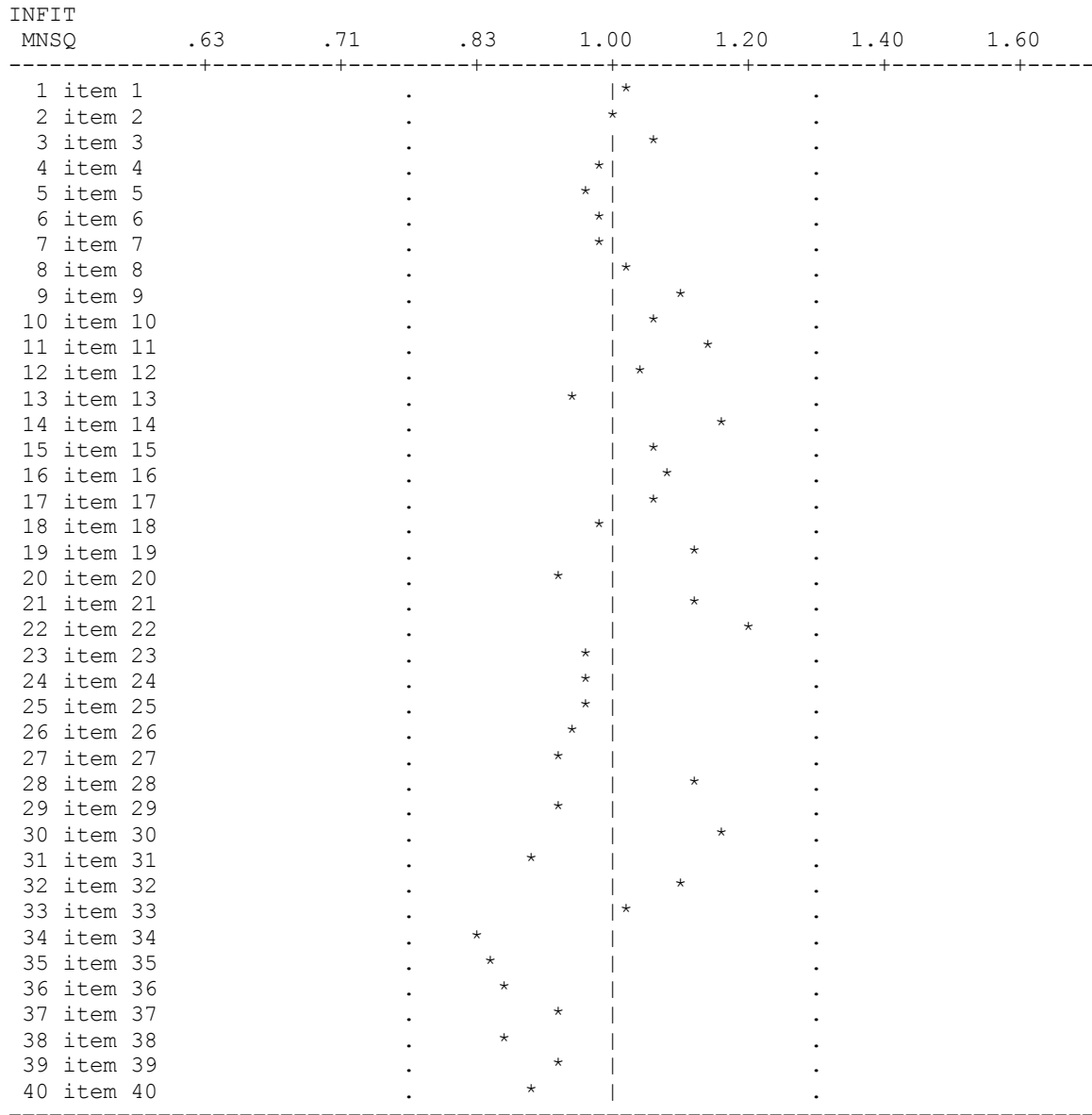
### 1. Karakteristik *Item Fit*

*PCM* merupakan pengembangan dari model Rasch, sehingga penentuan karakteristiknya cenderung mengacu kriteria pada Rasch. Menurut Adams & Khoo (1996); Didik Setyawarno (2017), bahwa dalam penggunaan program Quest untuk menetapkan fit item secara keseluruhan terhadap model berdasarkan besarnya nilai rata-rata *INFIT MNSQ* atau *INFIT Mean of Square* beserta standard deviasinya, atau

menggunakan nilai rata-rata *INFIT Mean of t* atau *INFIT t*. Demikian juga, untuk menetapkan *fit* tiap item terhadap model mengacu pada besarnya nilai *INFIT MNSQ* atau nilai *INFIT t* dari item, terlihat pada Gambar 1.

Pengujian item penelitian fkip-2022 (120)

Item Fit 20/ 9/22 6:16  
 all on all (N = 120 L = 40 Probability Level= .50)



Gambar 1. Analisis *Fit Items*

Hasil analisis memperlihatkan bahwa keseluruhan butir tes Geometri dalam penskalaan *PCM* berada di dalam interval  $0.70 \leq \text{INFIT MNSQ} \leq 1.33$ ; yang berarti keseluruhan butir soal adalah *fit* terhadap model pengukuran yang digunakan.

## 2. Karakteristik *Person Fit*

Demikian juga, untuk menetapkan *fit* person atau *fit case* terhadap model dengan menggunakan *Quest* mengacu kepada besarnya nilai *INFIT MNSQ* beserta standar

deviasinya. Ketentuan secara numerik atas *INFIT MNSQ* mengikuti kriteria untuk model Rasch (Adams & Khoo, 1996; Didik Setyawarno, 2017). Hasil analisis terlihat pada Gambar 2 berikut.

Pengujian item penelitian fkip-2022(120)

Case Estimates In input Order 20/ 9/22 6:16  
all on all (N = 120 L = 40 Probability Level= .50)

| NAME                  | SCORE     | MAXSCR     | ESTIMATE   | ERROR      | INFIT MNSQ  | OUTFT MNSQ  | INFIT t    | OUTFT t    |
|-----------------------|-----------|------------|------------|------------|-------------|-------------|------------|------------|
| 1 Aidi                | 94        | 154        | .57        | .15        | 1.30        | 1.38        | 1.71       | 1.56       |
| 2 Fathur Fild         | 89        | 154        | .47        | .15        | .71         | .72         | -1.90      | -1.25      |
| 3 h Zahrotu           | 84        | 154        | .36        | .15        | .70         | .75         | -2.00      | -1.10      |
| 4 Dharmawa            | 100       | 154        | .70        | .15        | .76         | .75         | -1.46      | -1.08      |
| 5 n                   | 88        | 154        | .44        | .15        | 1.11        | 1.09        | .70        | .46        |
| 6 fa Kusnaya          | 108       | 154        | .89        | .16        | 1.03        | 1.06        | .24        | .34        |
| 7 Aura Ang l          | 103       | 154        | .77        | .15        | .75         | .78         | -1.46      | -.94       |
| 8 Wijaya              | 86        | 154        | .40        | .15        | .83         | .83         | -1.03      | -.69       |
| 9 Aulia               | 78        | 154        | .23        | .15        | 1.13        | 1.18        | .78        | .79        |
| 10 Aloef Maula        | 107       | 154        | .86        | .15        | .81         | .81         | -1.02      | -.78       |
| 11 na Junirwin        | 111       | 154        | .96        | .16        | 1.26        | 1.32        | 1.30       | 1.23       |
| 12 dul Rasya          | 102       | 154        | .75        | .15        | 1.07        | 1.06        | .42        | .32        |
| 13 ikal               | 87        | 154        | .42        | .15        | 1.52        | 1.65        | 2.81       | 2.40       |
| 14 sady               | 97        | 154        | .64        | .15        | 1.04        | 1.10        | .30        | .49        |
| 15 llino Enj          | 92        | 154        | .53        | .15        | 1.07        | 1.09        | .45        | .45        |
| 16 Arjuna             | 85        | 154        | .38        | .15        | .72         | .77         | -1.80      | -1.01      |
| 17 Shafa Pur          | 86        | 154        | .40        | .15        | 1.22        | 1.24        | 1.30       | 1.03       |
| 18 Shafira            | 80        | 154        | .27        | .15        | 1.09        | 1.13        | .59        | .62        |
| 19 a Zanzabila        | 95        | 154        | .59        | .15        | 1.04        | 1.05        | .32        | .28        |
| 20 lla Leurila        | 95        | 154        | .59        | .15        | .65         | .68         | -2.38      | -1.48      |
| 21 usbah Elvr         | 93        | 154        | .55        | .15        | .75         | .75         | -1.59      | -1.13      |
| 22 waddah Wuli        | 121       | 154        | 1.23       | .17        | 1.13        | 1.10        | .62        | .42        |
| 23 Demas Dint         | 83        | 154        | .34        | .15        | 1.18        | 1.26        | 1.06       | 1.08       |
| 24 Adithya            | 89        | 154        | .47        | .15        | 1.20        | 1.24        | 1.18       | 1.03       |
| <b>25 ddin Alamly</b> | <b>91</b> | <b>154</b> | <b>.51</b> | <b>.15</b> | <b>1.07</b> | <b>1.12</b> | <b>.47</b> | <b>.55</b> |
| 113 utri Lestir       | 97        | 154        | .64        | .15        | 1.01        | .98         | .12        | -.03       |
| 114 sio Renold        | 87        | 154        | .42        | .15        | 1.41        | 1.43        | 2.30       | 1.70       |
| 115 il A.             | 115       | 154        | 1.06       | .16        | 1.10        | 1.09        | .55        | .41        |
| 116 a B.              | 82        | 154        | .32        | .15        | .98         | 1.00        | -.09       | .09        |
| 117 evin              | 109       | 154        | .91        | .16        | .99         | .99         | -.01       | .05        |
| 118 am                | 102       | 154        | .75        | .15        | 1.05        | .97         | .35        | -.07       |
| 119 diansyah          | 99        | 154        | .68        | .15        | 1.08        | 1.12        | .53        | .58        |
| 120 ya E.             | 99        | 154        | .68        | .15        | .87         | .84         | -.73       | -.65       |
| Mean                  |           |            | .65        |            | 1.02        | 1.01        | -.07       | -.03       |
| SD                    |           |            | .27        |            | .32         | .33         | 1.68       | 1.30       |

**Gambar 2. Analisis Person Fit**

Berdasarkan hasil analisis *person fit*, diperoleh bahwa dari 120 person yang memenuhi kriteria sebagai person yang dalam kondisi fit terhadap model terdapat 19 person yang berada di luar kriteria; yaitu person nomor 13, 20, 27, 28, 32, 33, 36, 41, 46, 47, 86, 88, 90, 92, 95, 96, 99, 103, dan 112. Berkenaan dengan ukuran sampel, Keeves & Masters (1999), mengungkapkan bahwa penentuan ukuran sampel pada model *PCM* dapat menggunakan sampel lebih kecil daripada ukuran sampel pada model *2-PL* atau *3-PL*. Dengan kondisi demikian, ke 19 person perlu dipertimbangkan keikutsertanya dalam analisis butir pada tahapan berikutnya. Apabila ukuran sampel masih memungkinkan, maka mereka dipertimbangkan untuk tidak diikuti pada langkah lebih lanjut.

### 3. Tingkat Kesukaran Butir

Tingkat kesukaran sebagai salah satu karakteristik butir soal yang memberikan informasi penting terkait kedudukan antar item dalam suatu perangkat tes, yakni butir soal itu berada dalam kategori mudah, sedang, sukar, sangat mudah, atau sangat sukar. Hasil analisisnya terlihat pada Gambar 3 berikut.

Pengujian item penelitian fkip-2022(122)

Item Estimates (Thresholds) In input Order 5/10/22 6:55  
all on all (N = 122 L = 40 Probability Level= .50)

| ITEM NAME  | SCORE MAXSCR |     | THRESHOLD/S   |             |            |             | INFT<br>MNSQ | OUTFT<br>MNSQ |
|------------|--------------|-----|---------------|-------------|------------|-------------|--------------|---------------|
|            |              |     | 1             | 2           | 3          | 4           |              |               |
| 5 item 5   | 317          | 488 | -1.94<br>.75  | -.11<br>.36 | .43<br>.31 | 1.61<br>.32 | 1.00         | 1.03          |
| 6 item 6   | 339          | 488 | -1.53<br>.70  | -.49<br>.45 | .29<br>.35 | 1.43<br>.30 | 1.02         | .99           |
| 7 item 7   | 289          | 488 | -1.88<br>.72  | -.43<br>.42 | .95<br>.32 | 1.59<br>.34 | 1.02         | 1.05          |
| 8 item 8   | 342          | 488 | -1.03<br>.53  | -.26<br>.39 | .23<br>.33 | 1.32<br>.31 | .98          | .98           |
| 9 item 9   | 313          | 488 | -2.69<br>1.06 | -.08<br>.33 | .49<br>.31 | 1.52<br>.32 | 1.08         | 1.08          |
| 10 item 10 | 276          | 488 | -2.31<br>.75  | .14<br>.33  | .88<br>.30 | 1.48<br>.31 | 1.07         | 1.13          |
| 11 item 11 | 274          | 488 | -3.09<br>1.06 | .24<br>.31  | .89<br>.30 | 1.39<br>.31 | 1.20         | 1.36          |
| 37 item 37 | 328          | 488 | -1.31<br>.53  | -.02<br>.34 | .51<br>.29 | 1.01<br>.27 | .90          | .89           |
| 38 item 38 | 355          | 488 | -1.16<br>.56  | -.32<br>.41 | .14<br>.36 | 1.21<br>.28 | .85          | .83           |
| 39 item 39 | 305          | 488 | -1.16<br>.50  | -.03<br>.34 | .57<br>.31 | 1.50<br>.30 | .94          | .92           |
| 40 item 40 | 327          | 488 | -1.06<br>.50  | -.05<br>.34 | .43<br>.31 | 1.17<br>.28 | .86          | .83           |
| Mean       |              |     | .00           |             |            |             | 1.00         | 1.02          |
| SD         |              |     | .17           |             |            |             | .10          | .14           |

**Gambar 3. Item Estimates**

Karakteristik butir soal jenis tingkat kesukaran pada item *PCM* bahwa tingkat kesukaran pada tahapan dibawahnya menuju kategori di atasnya tidak sama antar-item satu dan lainnya, sehingga koefisien delta untuk suatu tahapan kategori dibawahnya dan delta untuk tahapan kategori di atasnya adalah tidak sama antar-item yang satu dengan item-item berikutnya. Berdasarkan hasil analisis diperoleh bahwa terdapat butir soal kategorinya sangat mudah, mudah, sedang, sukar, dan sangat sukar. *PCM* memiliki tingkat kesukaran yang berjenjang, sesuai dengan banyaknya option. Jika option yang

disediakan sebanyak 4 atau  $m=4$ , maka banyaknya skor gradasi adalah  $m+1$ . Tingkat kesukaran yg tercakup dalam setiap butir soal adalah  $b_{j1}, b_{j2}, b_{j3}, b_{j4}$

#### 4. Reliabilitas

Menurut *IRT*, estimasi reliabilitas ditentukan berdasarkan item, (atau disebut indeks sparasi item) dan berdasarkan testi yang disebut indeks sparasi person. Estimasi reliabilitas berdasarkan testi (*case/person*) mempunyai kedudukan sama terhadap reliabilitas menuurt *CTT* (Bambang Subali & Pujiati Suyata, 2011). Dengan demikian reliabilitas perangkat tes Geometri ini sebesar 0.70. Hasil analisis (pada Gambar 4) menunjukkan hal tersebut.

```
Pengujian item penelitian fkip-2022(120)
-----
Case Estimates                               20/ 9/22 6:16
all on all (N = 120 L = 40 Probability Level= .50)
-----
Summary of case Estimates
=====
Mean                                     .65
SD                                       .27
SD (adjusted)                          .23
Reliability of estimate                 .70
```

**Gambar 4. Analisis Reliabilitas**

#### 5. Validitas

Karakteristik validitas isi pada perangkat tes dipenuhi pada waktu pelaksanaan penyusunan butir soal yang mendasarkan pada kompetensi Dasar, materi ajar, dan indikator soal yang dituangkan dalam kisi-kisi soal. Pengujian validitas item dapat ditunjukkan dengan kondisi *Fit Item* dan *fit testi* terhadap model (Bambang Subali & Pujiati Suyata, 2011). Item dikatakan valid apabila mengukur apa yang seharusnya untuk diukur. Suatu item disusun menurut model pengukuran yang dipilih, misal *PCM*, sehingga baik alat ukurnya atau individunya akan menentukan kondisi hasil pengukuran dengan *PCM* tersebut. Perpaduan item dan person terlihat pada analisis *Fit* dengan bantuan Quest. Menurut Adams & Khoo (1996), dalam program Quest, suatu item atau testi/case/person dinyatakan fit terhadap model apabila berada dalam kisaran *INFIT MNSQ* dari 0.77 sampai 1.33. Atau menggunakan kisaran harga  $t$ , yaitu  $\pm 2.0$  (sebagai pembulatan  $\pm 1.96$  jika  $\alpha=5\%$  (Keeves & Alagumulai, 1999). Dengan demikian, salah satu ciri item itu memiliki karakteristik validitas yang baik apabila *fit* terhadap model (Lihat Gambar 1).

#### KESIMPULAN

Pengujian butir soal geometri SMP dalam penskalaan *PCM* menghasilkan beberapa karakteristik butir, yang terdiri atas, *Fit item*, *Fit person*, Tingkat kesukaran, Reliabilitas, dan Validitas; yang masing-masing karakteristik tersebut belum menunjukkan hasil atau kondisi ideal. Beberapa kekurangan pada tiap-tiap karakter perlu adanya pengujian lanjutan, terutama dengan tidak melibatkan person-person yang memiliki bobot analisis Quest yang tidak ideal.

#### UCAPAN TERIMAKASIH

Terima kasih kepada Bapak/Ibu Kepala SMPN 2 Samarinda dan SMPN 8 Samarinda beserta guru matematika atas bantuan yang diberikan sehingga kegiatan penelitian ini terlaksana. Demikian juga, terima kasih kepada pihak-pihak yang telah membantu kegiatan penelitian ini.

**REFERENSI**

- Adams, R. J., & Khoo, S.-T. (1996). *QUEST: The interactive test analysis system*. Camberwell, VA: ACER Press.
- Allen, M. J & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Baker, J. G., Rounds, J. B., & Zevon, M. A. (2000). A comparison of graded response and Rasch partial credit model with subjective well-being. *Journal of Educational and Behavioral Statistics*, 25(3), 253–270.
- Bambang Subali, Kumaidi, & Nonoh Siti Aminah. (2018). *Pengembangan model asesmen konfirmatori penguasaan metode ilmiah pada mata pelajaran IPA di Sekolah Dasar*. ISBN: 978-602-498-019-1. Yogyakarta: UNY Press.
- Didik Setyowarno (September 2017). Materi kegiatan pelatihan Guru IPA SMP Sleman Yogyakarta oleh Universitas Negeri Yogyakarta dengan tema “Workshop Pemanfaatan Aplikasi Berbasis Komputer CMI-SIP SMA Sistem Informasi Penilaian Sekolah” bagi guru IPA Tingkat SMP di Sleman Yogyakarta “ di Laboratorium Komputer IPA, pada tanggal 9 September 2017.
- Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendikia Press.
- Djemari Mardapi. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Litera.
- Eccles, F. M. *An introduction to transformation geometry (Pengantar Geometri transformasi, alih bahasa Sudrajat, 2003)*. Bandung: Pustaka Setia.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologist*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fitzpatrick, A. R., Link, V. B., Yen, W. M., et al. (1996). Scaling performance assessment: A comparison of one-parameter and two-parameter partial credit models. *Journal of Educational Measurement*, 33(3), 291–314.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Keeves, J. P., & Alagumulai, S. (1999). New approaches to measurement. Dalam G. N. Masters & J. P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 23–42 ). Amsterdam: Pergamon.
- Kusaeri & Suprananto. (2012). *Pengukuran dan penilaian pendidikan*. Yogyakarta: Graha Ilmu.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- Reynold, C. R., Livingstone, R. B., & Willson, V. (2009). *Measurement and assessment in education*. New Jersey: Pearson Education, Inc.
- Saifuddin Azwar. (1999). *Penyusunan skala psikologi*. Yogyakarta: Pustaka Pelajar.
- Saifuddin Azwar. (2012). *Reliabilitas dan validitas*, (Edisi 4). Yogyakarta: Pustaka Pelajar.
- Sumadi Suryabrata. (2000). *Pengembangan alat ukur psikologis*. Yogyakarta: Andi.
- Sykes, R. C., & Yen, W. M. (2000). The scaling of mixed-item-format test with the one-parameter and two-parameter partial credit models. *Journal of Educational Measurement*, 37(3), 221–244.