

## Validitas Dan Reliabilitas Instrumen Penilaian Matematika Dengan Problem Solving Berbasis Penskoran Politomus

Dyah Ayu Fitriana

Sekolah Tinggi Agama Islam Khozinatul Ulum Blora,  
Jl. Mr. Iskandar No.42 Mlangsen, Kec. Blora Kota, Kabupaten Blora  
Email: dyahayufitriana@gmail.com

### Abstract

A single paragraph of about 250 words maximum. For research articles, abstracts should give a pertinent overview of the work. We strongly encourage authors to use the following style of structured abstracts, but without headings: (1) Background: Place the question addressed in a broad context and highlight the purpose of the study; (2) Purpose of the Study: Identify the purpose and objective of the study; (3) Methods: Describe briefly the main methods or theoretical framework applied; (4) Results: Summarize the article's main findings; and (5) Conclusions: Indicate the main conclusions or interpretations.

**Keywords:** *keyword 1; keyword 2; keyword 3 (List three to five pertinent keywords specific to the article; yet reasonably common within the subject discipline; use lower case except for names)*

### Abstrak

Penelitian ini merupakan penelitian pengembangan instrumen penilaian berbasis penskoran politomus pada pembelajaran matematika. Tujuan dari penelitian adalah untuk mengungkapkan validitas isi, validitas empiris, dan reliabilitas instrumen penilaian berbasis penskoran politomus yang telah disusun sebelumnya. Instrumen yang telah dikembangkan berupa instrumen tes uraian. Analisis validitas isi dari instrumen tes dianalisis dengan menggunakan CVI (Content Validity Index) oleh Aiken. Validitas empiris reliabilitas instrumen tes diestimasi dengan IRT (Item Response Theory). Reliabilitas instrumen tes dilihat dari klasik menggunakan Alpha Cronbach dan secara modern menggunakan fungsi informasi. Hasil dari penelitian ini menunjukkan bahwa (1) Instrumen penilaian terbukti valid dengan koefisien V Aiken 0,86 dengan grafik ICC yang Excellent. (2) Instrumen tes terbukti reliabel secara klasik dengan nilai 0,83 dan secara modern reliabel digunakan untuk kemampuan peserta didik dengan kategori sedang sampai dengan tinggi (-1,8 sampai dengan +1,8) dan INFIT MNSQ sesuai model PCM (Partial Credit Model). Berdasarkan hasil penelitian tersebut maka instrumen penilaian berbasis penskoran politomus layak digunakan.

**Kata kunci:** validitas isi, validitas empiris, instrumen penilaian, politomus

## A. Pendahuluan

Kegiatan penilaian dalam pendidikan merupakan hal yang wajib dilakukan di sekolah. Hal ini bertujuan untuk menilai hasil belajar peserta didik untuk dapat dilihat kemajuan belajar dan sebagai sarana perbaikan proses belajar. Jenis penilaian bisa berupa formatif, sumatif, diagnostik, penempatan, dan selektif. Prinsip penilaian yang dilakukan haruslah sesuai dengan Permendikbud No 23 Tahun 2016 tentang standar penilaian pendidikan, prinsip penilaian hasil belajar harus sah, objektif, dan adil.

Mencerdaskan kehidupan bangsa merupakan salah satu tujuan dari Indonesia. Tujuan tersebut terdapat dalam pembukaan UUD 1945 dan menjadi salah satu tujuan dari pendidikan nasional Indonesia yang diatur dalam UU Republik Indonesia Nomor 20 Tahun 2003. Oleh sebab itu, pendidikan yang diterapkan di Indonesia haruslah bertujuan untuk mencerdaskan warga negara Indonesia dalam hal ini adalah peserta didik agar mampu bersaing secara global.

Pendidikan yang diselenggarakan melalui proses pembelajaran di sekolah dapat menghasilkan lulusan yang mampu mengaplikasikan ilmu pengetahuan yang dimiliki untuk memecahkan masalah kehidupan sehari-hari (Badan Standar Nasional Pendidikan, 2006, p. 443). Berdasarkan acuan tersebut, kegiatan pembelajaran dapat didesain dengan memberikan tugas-tugas sebagai bentuk latihan untuk menyelesaikan permasalahan. Hal ini selaras dengan konsep literasi sains yang dikemukakan oleh PISA (Programme for International Student Assessment). Konteks literasi sains dalam penelitian ini dikhususkan hanya pada pembelajaran matematika.

Sains sebagai ilmu pengetahuan memiliki empat dimensi (Chiappetta & Koballa, 2010, p. 102) yaitu (1) science as a way of thinking, (2) science as a way of investigating, (3) science as a body of knowledge, dan (4) science and its interactions with technology and society. Keempat dimensi tersebut diperlukan dalam kegiatan pembelajaran sains.

Pembelajaran sains yang dilaksanakan harus dapat menjawab tantangan pada abad ke-21 yaitu untuk membentuk manusia yang dapat bertahan dengan perkembangan teknologi. Wagner (2008, pp. 14–38) merumuskan beberapa keterampilan yang dibutuhkan oleh manusia agar dapat bertahan hidup pada abad ke-21. Beberapa keterampilan tersebut adalah: (1) keterampilan berpikir kritis dan menyelesaikan masalah; (2) berkolaborasi dalam komunitas; (3) ketangkasan dan kemampuan beradaptasi; (4) berinisiasi dan jiwa bisnis; (5) kemampuan komunikasi secara lisan dan tulis; (6) mengakses dan menganalisis informasi; dan (7) imajinasi dan keingintahuan. Selain itu, literasi sains juga merupakan keterampilan yang penting pada abad ke-21 (P21, 2009, p. 2). Hal tersebut sejalan dengan hasil penelitian

Yore & Treagust (2006, p. 304) yang menunjukkan bahwa literasi sains dibutuhkan oleh individu pada abad 21 untuk dapat hidup dan memecahkan permasalahannya terkait dengan fenomena alam.

PISA merupakan program asesmen yang diadakan oleh OECD (Organization for Economic Co-operation and Development) untuk menilai kemampuan membaca, literasi matematika dan literasi sains peserta yang mengikuti tes tersebut. Pada tahun 2012 (OECD, 2014, p. 19) tes yang dilakukan oleh PISA diikuti oleh 510.000 peserta didik dari 65 negara. Indonesia berada pada ranking 64 dari 65 negara. Data tersebut menunjukkan bahwa kemampuan membaca, literasi matematika dan literasi sains peserta didik di Indonesia masih sangat rendah bila dibandingkan dengan 64 negara lain yang mengikuti program tersebut. Peringkat yang diperoleh Indonesia dalam PISA mengindikasikan bahwa peserta didik Indonesia memiliki literasi sains yang rendah. Selain itu, peringkat tersebut juga mengindikasikan bahwa Indonesia belum dapat bersaing dengan negara-negara lain.

Minimnya sumber pengetahuan juga berdampak pada ketidakpahaman mengenai penskoran yang sesuai dengan jenis tes yang digunakan agar dapat mencapai standar yang maksimal. Hal ini merupakan kekurangan yang didapat dalam suatu wawancara informal dengan beberapa guru di lingkungan Yogyakarta. Mencapai standar yang maksimal dibutuhkan pedoman penskoran yang tepat dalam melakukan kegiatan penilaian. Hasil belajar tidak terlepas dengan adanya instrumen penilaian yang digunakan sebagai alat evaluasi peserta didik. Penilaian berkeadilan tentunya diharapkan oleh segala pihak. Kenyataan, hal ini tidak pernah diperhatikan oleh para pendidik dalam melakukan suatu evaluasi.

Pendidikan berkeadilan dimaksud untuk dapat mengurangi bias maupun subyektivitas pendidik dalam mengevaluasi. Beberapa model respon butir selain dikotomus, yaitu model politomus. Model politomus pada teori respon butir antara lain Nominal Resons Model (NRM), Rating Scale Model (RCM), Partial Credit Modal (PCM), Graded Respon Model (GRM), dan Generalized Partial Credit Model (GPCM). GRM menilai banyaknya langkah yang benar dan memiliki indeks kesukaran yang terurut. PCM menilai banyaknya langkah yang benar pada kategori berjenjang, namun tingkat kesukaran tidak perlu terurut. Sehingga, dengan menggunakan penskoran politomus, butir-butir tes matematika dapat diskor menggunakan Partial Credit Modal (PCM), setiap langkah untuk menuju jawaban benar dihargai menggunakan penskoran ordinal (Heri, 2014). Selain itu, tingkat kesukaran dalam setiap langkah juga tidak perlu urut seperti halnya menemukan konsep dengan menghitung. Inilah yang dimaksudkan dengan prinsip penilaian berkeadilan. Berdasarkan observasi yang dilakukan di SMA Negeri di Kota Yogyakarta yang memiliki peringkat tinggi, sedang dan rendah bahwa asesmen yang digunakan oleh guru untuk mengukur kemampuan peserta didik adalah penilaian tes tertulis. Lebih lanjut diketahui pula bahwa penilaian tes tulis yang digunakan tidak disusun berdasarkan konsep pemecahan masalah dalam kehidupan sehari-hari. Hal tersebut belum mencerminkan tujuan pendidikan yang ditentukan oleh BSNP yang menuntut peserta didik untuk dapat mengaplikasikan pengetahuan yang

dimilikinya untuk memecahkan masalah dalam kehidupan sehari-hari dan penskoran yang digunakan masih belum adil bagi peserta didik.

Pemecahan masalah merupakan tujuan utama dari sebuah pembelajaran matematika. Untuk dapat mengembangkan kemampuan pemecahan masalah, dibutuhkan pengalaman yang lebih luas dan soal yang beragam kepada peserta didik. Penting untuk diketahui sebagai pendidik yaitu alasan penggunaan model pemecahan masalah. Masalah didefinisikan sebagai situasi yang membuat individu merasa tidak nyaman dan dibutuhkan solusi yang masuk akal untuk mencapai keseimbangan (Incebacak & Ersoy, 2016: 275). Dalam Posamentier & Krulik (2009: 2) mengemukakan bahwa pemecahan masalah dapat melatih cara berpikir peserta didik. Problem solving merupakan bagian dari setiap kehidupan (Gojak, 2011: 15).

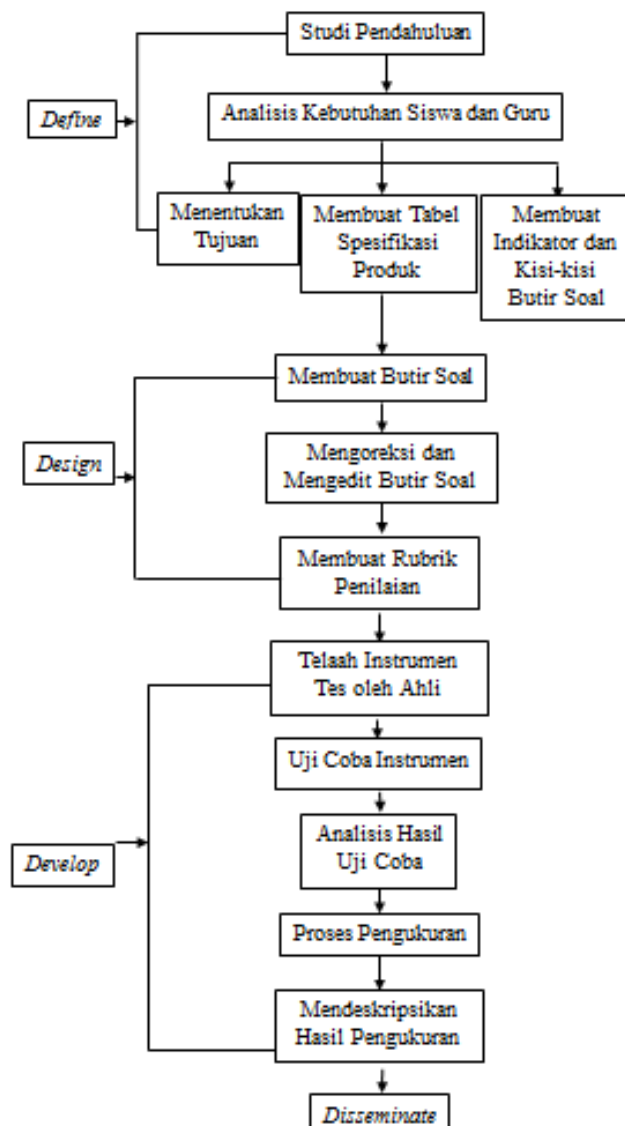
Selain dengan memperbaiki proses pembelajaran yang dilaksanakan, peningkatan kualitas pembelajaran dapat dilakukan melalui perbaikan pada instrumen penilaian yang digunakan untuk mengukur hasil belajar peserta didik. Perbaikan dalam aspek penilaian dapat dilakukan dengan memberikan panduan dan pelatihan dalam melakukan penilaian (Kartowagiran & Jaedun, 2016, p. 140). Asesmen yang digunakan untuk memperoleh, menyintesis dan menginterpretasikan informasi yang diperoleh dari peserta didik dengan tujuan untuk menarik keputusan tentang peserta didik di dalam kelas dikenal dengan istilah classroom assessment. Classroom assessment yang digunakan untuk menilai kinerja peserta didik selama pembelajaran dikenal sebagai assessment for learning. Dalam pembelajaran, assessment for learning menyatu dengan rencana pelaksanaan pembelajaran yang disusun atau digunakan oleh guru.

Salah satu bentuk assessment for learning adalah penilaian kinerja. Penilaian kinerja digunakan untuk mengetahui kemampuan peserta didik dalam menerjemahkan pemahamannya menjadi suatu kerja nyata, sehingga penilaian dapat menunjukkan kemampuan peserta didik yang sesungguhnya.

Instrumen yang digunakan dalam penilaian haruslah memiliki syarat layak untuk digunakan yaitu valid dan reliabel (Mardapi, 2012, pp. 37–96; Sumintono & Widhiarso, 2015, pp. 7–12). Oleh karena itu, instrumen penilaian berbasis pemecahan masalah yang telah dikembangkan harus memenuhi kedua syarat tersebut. Oleh karena itu, uji validitas dan reliabilitas instrumen penilaian perlu dilakukan.

## **B. Metode Penelitian**

Penelitian ini merupakan bagian dari penelitian pengembangan instrumen penilaian matematika dengan problem solving berbasis penskoran politomus. Instrumen yang telah memenuhi standar valid dan reliabel dapat digunakan untuk tahap pengukuran. Tahapan penelitian secara keseluruhan yang disajikan pada Gambar 1 mengacu pada pembuatan instrumen tes Oriondo dan Antonio.



Gambar 1. Prosedur Pengembangan Penilaian

Model pengembangan Oriundo and Antonio memiliki langkah-langkah, yaitu 1) Planning the test, 2) Trying out the test, and 3) Establishing the Validity Test, 4) Establishing Test Reliability, and 5) Interpreting the Test Score. Langkah ini dipadukan dengan pengembangan 4D sesuai dengan Gambar 1, tujuan instrumen penilaian dapat mengukur kemampuan pemecahan masalah peserta didik pada pelajaran matematika berbasis penskoran politomus. Penelitian ini dilakukan dengan mengambil dua sampai tiga kelas XI IPA di SMA Negri se-Kota Yogyakarta dengan total subjek 510 peserta didik. Teknik analisis data menggunakan deskriptif kualitatif dan analisis kuantitatif. Kualitatif bertujuan untuk melihat konstruk instrumen melalui expert judgement. Analisis kuantitatif digunakan untuk melihat validitas dan reliabilitas instrumen. Validitas dilihat dari isi dan secara empirik. Dilihat secara isi melalui hasil penilaian expert judgement yang dihitung menggunakan V Aiken dimana menurut Retnawati (2017), indeks nilai aiken apabila kurang dari sama dengan 0,4 dikatakan validitasnya kurang, 0,4 hingga 0,8 dikatakan validitasnya sedang, dan diatas 0,8 dikatakan sangat valid.

### Validitas Isi

Validitas isi menggunakan formula V Aiken. Penilaian diberikan dengan cara memberikan angka 1 sampai 4 di tiap butir soal tes. Ketentuan pemberian angka dengan ketentuan 1= sangat tidak cocok, 2= tidak cocok, 3= cocok, 4= sangat cocok. Hasil telaah ahli kemudian diolah menggunakan formula, (Aiken, 1980: 956)

$$V = \frac{\sum s}{n(e - 1)}$$

#### Keterangan:

|    |                                               |
|----|-----------------------------------------------|
| V  | = indeks validitas V Aiken                    |
| s  | = $r - lo$                                    |
| r  | = angka yang diberikan oleh validator         |
| lo | = angka penilaian terendah                    |
| n  | = jumlah rater                                |
| e  | = banyaknya kategori yang dapat dipilih rater |

Analisis menggunakan formula V Aiken selanjutnya diinterpretasikan dengan ketentuan indeks V Aiken memiliki nilai validitas yang kurang, indeks V Aiken memiliki nilai validitas yang sedang, dan indeks V Aiken memiliki nilai validitas yang baik (Retnawati, 2016: 19).

### Validitas Konstruk

Validitas konstruk menggunakan uji asumsi IRT yaitu 1) uji unidimensi; 2) uji asumsi independensi lokal; 3) invariansi parameter.

Uji asumsi unidimensi harus terpenuhi untuk membuktikan bahwa setiap butir soal hanya mengukur satu kemampuan saja. Cara menguji asumsi ini dengan menggunakan faktor analisis yang menghasilkan output KMO, eigen value, varians, dan komponen faktor. Uji asumsi unidimensi terpenuhi jika butir soal terbukti hanya mengukur satu dimensi yang dominan (Hambleton & Swaminathan, 1985: 16).

Uji asumsi independensi lokal merupakan syarat selanjutnya yang harus dipenuhi dalam analisis IRT. Asumsi ini bertujuan untuk melihat apakah kemampuan peserta didik independen terhadap butir soal. Artinya, jawaban peserta didik terhadap suatu butir soal tidak akan mempengaruhi jawaban terhadap butir soal yang lainnya.

Invariansi parameter membuktikan bahwa butir soal tidak memiliki ketergantungan pada distribusi parameter kemampuan dari peserta didik dan sebaliknya parameter kemampuan peserta didik tidak bergantung pada butir soal. Kemampuan peserta didik tidak akan berubah karena mengerjakan paket soal yang memiliki parameter butir soal yang

berbeda dan parameter butir soal juga tidak akan berubah meski diberikan kepada kelompok siswa yang berbeda. Ada dua invariansi parameter yaitu invariansi parameter butir soal dan invariansi parameter kemampuan.

### Reliabilitas

Estimasi nilai reliabilitas dari instrumen tersebut menggunakan formula Kuder-Richardson 20. Menurut Retnawati (2016), mengatakan bahwa besaran reliabilitas memiliki rentang angka -1,00 hingga +1,00. Semakin tinggi koefisien reliabilitas, maka instrumen tersebut memiliki reliabilitas yang tinggi pula. Formula yang digunakan sebagai berikut

$$KR20 = \left[ \frac{k}{(k-1)} \right] \left[ \frac{s^2_x - \sum_i^j p_i^2 (1-p_i)}{s^2_x} \right]$$

Keterangan:

$k$  = jumlah butir soal dalam tes

$s^2_x$  = variansi skor total

$p_i$  = proporsi subjek yang menjawab benar pada satu butir soal

Selanjutnya, secara IRT diestimasi menggunakan kurva Total Info dan SEM. Kriteria batas reliabilitas instrumen ditentukan melalui titik potong kurva Total Info dan SEM tersebut.

Analisis butir soal pada penelitian ini dilakukan setelah peneliti melakukan uji coba produk dengan skala terbatas. Analisis uji coba instrumen dilakukan pada model 1PL dengan bantuan software QUEST untuk mengetahui tingkat kesukaran butir soal yang sudah dikembangkan dengan pedoman penskoran yang berbasis politomus partial credit model. Kemudian karakteristik yang diolah, diinterpretasikan berdasarkan kriteria tingkat kesulitan butir soal. Kriteria model 1PL memiliki karakteristik baik jika berada direntang nilai -2 hingga +2, selain itu maka butir soal mengalami revisi. Butir soal yang sudah sesuai, maka dapat digunakan untuk uji penelitian dalam hal ini skala yang besar

### C. Hasil dan Pembahasan

Validitas isi ditentukan berdasarkan judgement expert yang terdiri dari ahli, praktisi, dan peer review. Hasil expert judgement diolah menggunakan persamaan V Aiken yang diperoleh hasil antara rentang 0,7 – 0,8. Menurut Retnawati (2017), indeks nilai aiken apabila kurang dari sama dengan 0,4 dikatakan validitasnya kurang, 0,4 hingga 0,8 dikatakan validitasnya sedang, dan diatas 0,8 dikatakan sangat valid. Berdasarkan perhitungan validitas Aiken, butir soal utama sebanyak 16 soal utama dan 8 soal anchor untuk paket A dan B memiliki kriteria valid. Item tersebut sudah direvisi sesuai saran dari para ahli agar bisa digunakan. Sehingga, instrumen tersebut dapat digunakan untuk mengambil data agar memperoleh validitas dan reliabilitas secara empiris. Hasil perhitungan disajikan dalam tabel 1.

Tabel 1 Hasil Validitas Aiken

| Butir                                                     | Nilai Validitas Aiken dari<br>7 Rater |
|-----------------------------------------------------------|---------------------------------------|
| A1, A2, A3, A6, A8, 9, 11, 13,<br>B19, B20, B22, B23, B24 | 0,8                                   |
| A4, A7, 10, 12, 15, 16, B18                               | 0,73                                  |
| A5, 14                                                    | 0,7                                   |
| B21                                                       | 0,67                                  |

Validitas Konstruk (Uji Asumsi IRT)

1. Uji Asumsi Unidimensi

Hasil dari analisis faktor disajikan pada Tabel 2.

Tabel 2 Hasil Uji KMO dan Bartlett

|                                                  |      |           |
|--------------------------------------------------|------|-----------|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. |      | ,965      |
| Approx. Chi-Square                               |      | 20134,420 |
| Bartlett's Test of Sphericity                    | Df   | 276       |
|                                                  | Sig. | ,000      |

Berdasarkan Tabel 2, hasil KMO sebesar 0,965 (lebih dari 0,50). Artinya bahwa ukuran sampel yang digunakan pada uji coba sudah mencukupi. Nilai khi-khuadrat pada Uji Bartlett sebesar 20134,420 dengan derajat kebebasan 780 dan nilai-p lebih kecil dari 0,01. Matriks ini dapat dianalisis faktor jika nilai  $KMO > 0,5$ .

Uji asumsi unidimensi terpenuhi jika butir soal terbukti hanya mengukur satu dimensi yang dominan (Hambleton & Swaminathan, 1985: 16). Berdasarkan hasil analisis faktor terdapat 4 faktor yang memiliki nilai eigen value lebih dari 1. Tetapi faktor 1 merupakan faktor yang dominan karena memiliki nilai eigen sebesar 15,555 dan nilai tersebut lebih besar dibanding faktor yang lainnya. Faktor 1 juga memiliki nilai eigen lebih besar dari 2 kali faktor yang lainnya. Sehingga, instrumen penilaian kemampuan pemecahan masalah Matematika SMA bersifat unidimensi

2. Uji Asumsi Independensi Lokal

Uji asumsi ini dapat dilakukan dengan melihat matriks kovarian berdasarkan kemampuan peserta didik yang dikelompokkan menjadi beberapa kelompok. Asumsi ini terpenuhi jika nilai kovarian antar interval kemampuan kecil (mendekati nol). Adapun matriks kovarian disajikan pada Tabel 3

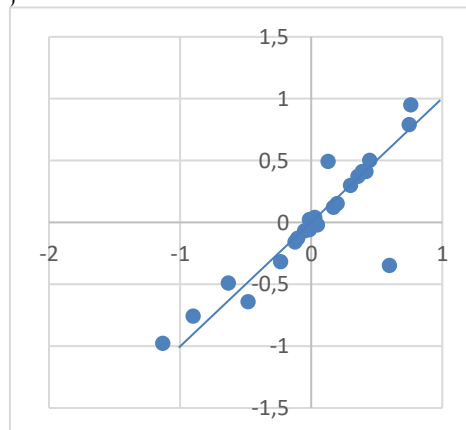
**Tabel 3 Matriks Kovarian Kemampuan Peserta Didik**

|     | K1      | K2      | K3      | K4      | K5      | K6      | K7      | K8     | K9     | K10    |
|-----|---------|---------|---------|---------|---------|---------|---------|--------|--------|--------|
| K1  | 0,0006  |         |         |         |         |         |         |        |        |        |
| K2  | -0,0001 | 0,0014  |         |         |         |         |         |        |        |        |
| K3  | 0,0000  | 0,0000  | 0,0020  |         |         |         |         |        |        |        |
| K4  | 0,0006  | -0,0007 | 0,0000  | 0,0024  |         |         |         |        |        |        |
| K5  | -0,0004 | -0,0010 | -0,0002 | 0,0010  | 0,0047  |         |         |        |        |        |
| K6  | -0,0003 | 0,0006  | -0,0003 | -0,0006 | 0,0001  | 0,0024  |         |        |        |        |
| K7  | -0,0001 | 0,0004  | 0,0001  | -0,0008 | -0,0015 | 0,0000  | 0,0014  |        |        |        |
| K8  | 0,0001  | -0,0001 | 0,0003  | 0,0004  | 0,0003  | 0,0002  | -0,0002 | 0,0008 |        |        |
| K9  | 0,0002  | -0,0004 | -0,0001 | 0,0010  | 0,0006  | -0,0002 | -0,0003 | 0,0002 | 0,0013 |        |
| K10 | -0,0001 | -0,0007 | 0,0000  | 0,0008  | 0,0018  | 0,0000  | -0,0005 | 0,0002 | 0,0006 | 0,0014 |

Berdasarkan Tabel 3, nilai variansi kovarian antar kelompok interval kemampuan peserta didik yang membentuk garis diagonal memiliki nilai kecil hampir mendekati nol. Sehingga, karena tidak adanya korelasi maka asumsi independensi lokal terpenuhi.

### 3. Invariansi Parameter

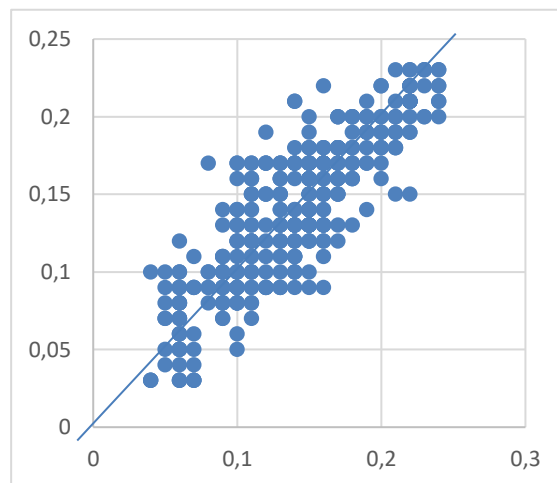
Asumsi invariansi parameter butir terpenuhi apabila korelasinya positif dan tinggi. Invariansi parameter butir disajikan dalam Gambar 2



Gambar 2 Scree Plot Invariansi Parameter Tingkat Kesulitan Soal

Berdasarkan Gambar 2, terlihat nilai estimasi terletak mengikuti garis linier. Hasil korelasi tingkat kesulitan butir dari kelompok peserta didik ganjil dan genap termasuk dalam kategori tinggi sebesar 0,884066 sehingga, invariansi tingkat kesulitan soal terpenuhi.

Invariansi yang kedua yaitu kemampuan peserta tes dengan cara membagi kelompok menjadi 2 kelompok. Hasil estimasi parameter kemampuan dari masing-masing kelompok dibuat scree plot dan dikorelasikan. Invariansi kemampuan disajikan dalam Gambar 5.



Gambar 3 Scree Plot Invariansi Kemampuan

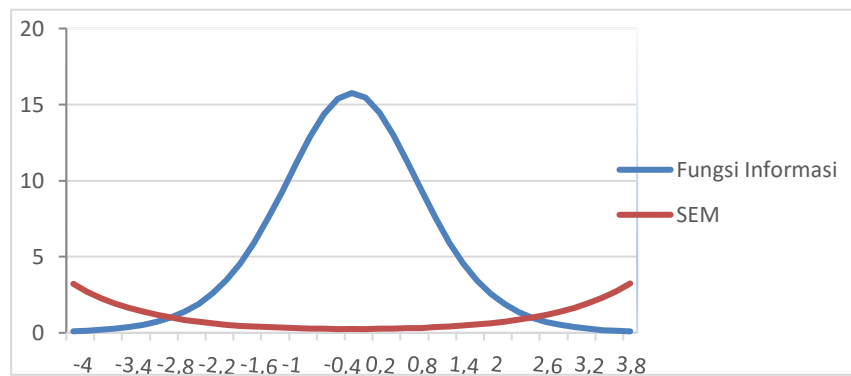
Berdasarkan Gambar 3, hasil scree plot menunjukkan bahwa nilai-nilai estimasi mengikuti garis linier. Nilai korelasi sebesar 0,796 masuk dalam kategori tinggi. Sehingga, invariansi kemampuan terpenuhi. Gambar 2 dan Gambar 3 dapat ditarik kesimpulan bahwa, instrumen ini telah memenuhi asumsi invariansi.

### Reliabilitas

Hasil estimasi reliabilitas secara klasik menggunakan Alpha Cronbach diperoleh nilai 0,82 (lihat tabel 4) . Berdasarkan kategori Triton (2006), instrumen ini sangat reliabel. Sehingga bisa digunakan sebagai instrumen alat ukur. Hasil analisis menggunakan pendekatan IRT (Item Respon Theory) menggunakan fungsi informasi dan total info diperoleh bahwa instrumen ini reliabel apabila digunakan untuk mengukur antara -1,8 – 1,8 (lihat gambar 4).

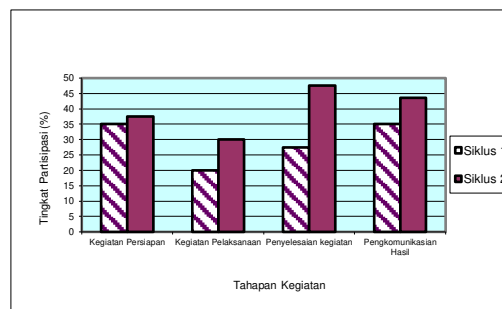
Tabel 4 Hasil Estimasi *Item Soal* dan *Person*

| No | Uraian                                                    | Estimasi untuk <i>Item</i> | Estimasi untuk <i>Person</i> |
|----|-----------------------------------------------------------|----------------------------|------------------------------|
| 1  | Nilai rata-rata dan simpangan baku                        | $-0,11 \pm 0,32$           | $0,09 \pm 0,05$              |
| 2  | Nilai rata-rata dan simpangan baku yang sudah disesuaikan | $0,00 \pm 0,29$            | $-0,11 \pm 0,00$             |
| 3  | Reliabilitas                                              | 0,82                       |                              |
| 4  | Nilai rata-rata dan simpanan baku INFIT MNSQ              | $1,00 \pm 0,07$            | $1,02 \pm 0,19$              |
| 5  | Nilai rata-rata dan simpanan baku OUTFIT MNSQ             | $0,99 \pm 0,07$            | $0,99 \pm 0,21$              |
| 6  | Nilai rata-rata dan simpangan baku INFIT t                | $0,30 \pm 2,10$            | $0,18 \pm 1,22$              |
| 7  | Nilai rata-rata dan simpangan baku OUTFIT t               | $-0,03 \pm 1,14$           | $0,01 \pm 0,74$              |



Gambar 4. Fungsi Informasi dan SEM

berupa gambar, atau data yang dibuat gambar/skema/grafik/diagram/sebangsa-nya, pemaparannya juga mengikuti aturan yang ada; judul atau nama gambar ditaruh di bawah gambar, dari kiri, dan diberi jarak 1 spasi (*at least 12*) dari gambar. Bila lebih dari 1 baris, antarbaris diberi spasi tunggal, atau *at least 12*. Sebagai contoh, dapat dilihat pada Gambar 1. berikut.



Gambar 1. Judul gambar tabel .....

#### D. Kesimpulan

Hasil penelitian menunjukkan bahwa 1) Instrumen penilaian matematika dengan problem solving berbasis penskoran politomus memenuhi persyaratan valid berdasarkan expert judgement yang dianalisis menggunakan Validitas Aiken, dan valid berdasarkan konstruk menggunakan asumsi IRT. 2) Instrumen penilaian matematika dengan problem solving berbasis penskoran politomus memenuhi persyaratan reliabel berdasarkan klasik yang diestimasi menggunakan Alpha Cronbach dan Modern yang diestimasi menggunakan Fungsi Informasi dan SEM. Sehingga, instrumen penilaian matematika dengan problem solving berbasis penskoran politomus dapat digunakan untuk praktisi dan akademisi untuk mengetahui tingkat kemampuan pemecahan masalah matematika SMA kelas XI IPA

#### Daftar Pustaka

- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142.  
<https://doi.org/10.1177/0013164485451012>
- Azwar, S. (2012). *Reliabilitas dan validitas* (4th ed.). Yogyakarta: Pustaka Pelajar.

- Azwar, S. (2015). *Metode penelitian*. Yogyakarta: Pustaka Pelajar.
- Badan Standar Nasional Pendidikan. (2006). *Panduan penyusunan kurikulum tingkat satuan pendidikan jenjang pendidikan dasar dan menengah*. Jakarta: BSNP. Retrieved from [http://bsnp-indonesia.org/id/wp-content/uploads/kompetensi/PanduanUmum\\_KTSP.pdf](http://bsnp-indonesia.org/id/wp-content/uploads/kompetensi/PanduanUmum_KTSP.pdf)
- Bashooir, K., & Supahar, S. (2016). Analisis aspek kinerja literasi sains pada materi kalor Fisika. *UPEJ Unnes Physics Education Journal*, 5(1). Retrieved from <https://journal.unnes.ac.id/sju/index.php/upej/article/view/12711>
- Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. *School Science and Mathematics*, 112(1), 3–11. <https://doi.org/10.1111/j.1949-8594.2011.00109.x>
- Chiappetta, E. L., & Koballa, T. R. (2010). *Science instruction in the middle and secondary schools: developing fundamental knowledge and skills* (7th ed.). USA: Pearson Education, Inc.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Departemen Pendidikan dan Kebudayaan. (2001). *Kamus besar bahasa Indonesia* (3rd ed.). Jakarta: Balai Pustaka.
- Department of Education. (2009). *Report of the STEM review*. Retrieved from [https://www.education-ni.gov.uk/sites/default/files/publications/de/Report of the STEM Review 2009\\_1.PDF](https://www.education-ni.gov.uk/sites/default/files/publications/de/Report%20of%20the%20STEM%20Review%202009_1.PDF)
- Gonzalez, H. B., & Kuenzi, J. J. (2012). Science, technology, engineering, and mathematics (STEM) education: a primer. Retrieved from <https://fas.org/sgp/crs/misc/R42642.pdf>
- Hernandez, P. R., Bodin, R., Elliott, J. W., Ibrahim, B., Rambo-Hernandez, K. E., Chen, T. W., & de Miranda, M. A. (2014). Connecting the STEM dots: measuring the effect of an integrated engineering design intervention. *International Journal of Technology and Design Education*, 24(1), 107–120. <https://doi.org/10.1007/s10798-013-9241-0>
- Ismail, I., Permanasari, A., & Setiawan, W. (2016). Efektivitas virtual lab berbasis STEM dalam meningkatkan literasi sains siswa dengan perbedaan gender. *Jurnal Inovasi Pendidikan IPA*, 2(2), 190. <https://doi.org/10.21831/jipi.v2i2.8570>
- Kartowagiran, B., & Jaedun, A. (2016). Model asesmen autentik untuk menilai hasil belajar siswa sekolah menengah pertama (SMP): implementasi asesmen autentik di SMP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(2), 131. <https://doi.org/10.21831/pep.v20i2.10063>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Mardapi, D. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha

Medika.

- OECD. (2014). PISA 2012 results: what students know and can do student performance in mathematics, reading and science volume I. Paris: OECD Publishing.
- P21. (2009). 21st century skills map. Retrieved from [http://www.p21.org/storage/documents/21st\\_century\\_skills\\_english\\_map.pdf](http://www.p21.org/storage/documents/21st_century_skills_english_map.pdf)
- Presiden Republik Indonesia. Undang- Undang Republik Indonesia nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional (2003). Indonesia.
- Reeve, E. M. (2013). Implementing science, technology, mathematics, and engineering (STEM) education in Thailand and in ASEAN. Retrieved from [http://dpst-apply.ipst.ac.th/specialproject/images/IPST\\_Global/document/Implementing STEM in ASEAN - IPST May 7 2013 - Final.pdf](http://dpst-apply.ipst.ac.th/specialproject/images/IPST_Global/document/Implementing%20STEM%20in%20ASEAN%20-%20IPST%20May%207%202013%20-%20Final.pdf)
- Retnawati, H. (2016). Validitas reliabilitas dan karakteristik butir. Yogyakarta: Parama Publishing.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18839484>
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103. [https://doi.org/10.1207/S15327752JPA8001\\_18](https://doi.org/10.1207/S15327752JPA8001_18)
- Subali, B., & Suyata, P. (2012). Pengembangan item tes konvergen dan divergen: penyelidikan validitasnya Secara empiris. Yogyakarta: Diandra Pustaka Indonesia.
- Sumintono, B., & Widhiarso, W. (2015). Aplikasi pemodelan RASCH pada assessment pendidikan. Cimahi: Tim Komunikata Publishing House.
- Supahar, & Prasetyo, Z. K. (2015). Pengembangan instrumen penilaian kinerja kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(1), 96–108. Retrieved from <https://journal.uny.ac.id/index.php/jpe/article/view/4560>
- Supahar, S. (2014). The estimation of inquiry performance test items of high school physics subject with quest program. In *International Conference on Research, Implementation And Education of Mathematics And Sciences*. Yogyakarta: Yogyakarta State University.
- Supahar, S. (2015). Applying content validity ratios (CVR) to the quantitative content validity of physics learning achievement tests. In *International Conference on Research, Implementation And Education of Mathematics And Sciences*. Yogyakarta: Yogyakarta State University.
- Wagner, T. (2008). *The global achievement gap*. New York: Basic Book.
- Yore, L. D., & Treagust, D. F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. *International Journal of Science Education*, 28(2–3), 291–314. <https://doi.org/10.1080/09500690500336973>