

Volume 4 Nomor 1, Februari 2019, halaman 9-20

**PENGEMBANGAN INSTRUMEN ASESMEN OTENTIK TES
KETERAMPILAN BERPIKIR KRITIS PADA PEMBELAJARAN
MATEMATIKA**

***DEVELOPMENT OF AUTHENTIC ASSESSMENT INSTRUMENTS OF
CRITICAL THINKING SKILLS TEST ON MATHEMATICS LEARNING***

Rivo Panji Yudha

Universitas 17 Agustus 1945 Cirebon, Jln. Perjuangan No.17 Kota Cirebon,
rivoyudha@yahoo.co.id

ABSTRAK

Penelitian ini bertujuan untuk menghasilkan instrumen asesmen otentik yang layak digunakan untuk mengukur keterampilan berpikir kritis peserta didik. Desain Penelitian yang digunakan adalah penelitian pengembangan modifikasi model Wilson dan Oriondo dan Antonio. Validitas isi instrumen dibuktikan dengan *expert judgement teknik Delphi* oleh ahli dan praktisi. Validitas empiris dibuktikan dan dianalisis menggunakan program Winsteps 3.73. Data politomus hasil uji coba dianalisis menggunakan model *Rasch* dengan bantuan program Winsteps 3.73. Uji coba instrumen melibatkan 30 peserta didik kelas XI AK SMK Wahidin Cirebon. Hasil penelitian menunjukkan bahwa telah dihasilkan instrumen penilaian otentik yang layak untuk mengukur berpikir kritis peserta didik. Karakteristik instrumen penilaian otentik ditinjau dari validitas isi oleh ahli dan praktisi termasuk dalam kategori sangat baik, sedangkan untuk validasi empirisnya berdasarkan tiga kriteria yaitu *OUTFIT MNSQ*, *ZSTD*, *Pt Mean Corr* untuk 10 butir instrumen keterampilan berpikir kritis. Reliabilitas instrumen keterampilan berpikir kritis berdasarkan *index sparation item* sangat bagus dengan nilai $>0,90$, dan kualitas implementasi asesmen otentik keterampilan berpikir kritis pada pembelajaran matematika adalah praktis.

Kata Kunci: Instrumen, Asesmen, Keterampilan Berpikir Kritis

ABSTRACT

*This research supports producing authentic assessment instruments that are suitable to be used to measure the intelligence of students. The research design used was a research on the development of modified models of Wilson and Oriondo and Antonio. The content validity of the instrument is proven by expert judgment. Delphi techniques by experts and prove. Empirical validity is proven and analyzed using the Winsteps 3.73 program. The results of the experimental polytomus try to use the Rasch model with the help of the Winsteps program 3.73. The trial of the participation instruments of 30 students of class XI AK Wahidin Cirebon. The results of the study show that it has produced an instrument that proves feasible to determine that the participants are critical students. The characteristics of the approved instruments in terms of content validity by experts and evaluations are included in a very good category, while empirical validation is based on three criteria namely *MNSQ OUTFIT*, *ZSTD*, *Pt Mean Corr* for 10 instrument items that require truth assistance. Reliability of critical assessment instruments based on the index of spatial*

items is very good with values > 0.90, and the quality of implementing authentic assessments of critical thinking skills in mathematics learning is practical.

Keywords: *Instruments, Assessment, Critical Thinking Skills*

How to Cite: Yudha, R.P. (2019). Pengembangan Instrumen Asesmen Otentik Tes Keterampilan Berpikir Kritis pada Pembelajaran Matematika. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, Vol.4, No.1, 9-20.

PENDAHULUAN

Berbagai masalah yang dihadapi dalam meningkatkan kualitas pendidikan di Indonesia, dimulai dari tingkat dasar pendidikan sampai ke perguruan tinggi. Proses belajar mengajar tidak hanya berbicara tentang proses, tetapi juga berbicara tentang hasilnya. Karenanya, untuk mengetahui hasil dari proses itu, guru atau dosen harus menggunakan tes sebagai alat untuk mengukur kemampuan siswa atau kinerja, dan memutuskan, apakah siswa dapat lulus atau tidak. Dalam proses belajar mengajar, dosen tidak hanya fokus pada proses pengajaran, tetapi juga pada bagaimana mereka mengukur hasil siswa atau magang mereka. Penilaian untuk pembelajaran adalah perspektif baru tentang sistem penilaian dalam pendidikan. Praktek tradisional adalah untuk mengevaluasi hasil adalah Penilaian Pembelajaran.

Kurikulum 2013 saat ini menekankan peserta didik untuk berperan lebih aktif dalam proses pembelajaran sehingga dalam melakukan penilaiannya juga mengalami perkembangan yaitu dari penilaian hasil menjadi penilaian proses yang mempertimbangkan sikap, perilaku, dan akhlak sebagai bagian tak terpisahkan ketika melakukan penilaian. Penilaian otentik (*authentic assessment*) memiliki relevansi kuat terhadap pendekatan ilmiah dalam pembelajaran sesuai dengan tuntutan Kurikulum 2013. Peraturan Menteri Nomor 22 tahun 2016 menghendaki bahwa penilaian proses pembelajaran menggunakan pendekatan penilaian otentik (*authentic assessment*) yang menilai kesiapan peserta didik, proses, dan hasil belajar secara utuh. Penilaian otentik (*authentic assessment*) adalah pengukuran yang bermakna atas hasil belajar peserta didik untuk ranah sikap, keterampilan, dan pengetahuan.

Pada kurikulum 2013 keterampilan proses sains dan kemampuan berpikir tingkat tinggi merupakan dua hal yang tidak kalah penting. Keterampilan proses merupakan salah satu aspek dari matematika. Keterampilan proses digunakan untuk membangun tubuh pengetahuan yang merupakan kumpulan dari ilmu pengetahuan. Selain keterampilan proses, kemampuan berpikir tingkat tinggi yaitu berpikir kritis perlu dimiliki peserta didik

karena dapat membantu peserta didik untuk mengembangkan potensi intelektualnya, sebagaimana dijelaskan oleh Johnson (2007) yang menyatakan bahwa kemampuan berpikir kritis diperlukan bagi seseorang, hal ini dianggap penting karena mampu membantu peserta didik mengembangkan potensi intelektualnya, memiliki kemampuan untuk mengevaluasi secara sistematis, dan mampu berpendapat secara terorganisasi.

Pada implementasinya, khususnya pembelajaran matematika masih sering dijumpai soal-soal evaluasi pada penilaian formatif yang cenderung menuntut kemampuan ingatan dan memecahkan masalah lewat solusi matematis dalam bentuk soal pilihan ganda yang sering dikontradiksikan dengan penilaian otentik. Bentuk soal pilihan ganda dipilih karena model ini memudahkan guru dalam membuat sebaran soal terkait dengan materi yang dipelajari, selain itu juga memudahkan guru dalam melakukan penilaian karena pada bentuk soal pilihan ganda hanya terdapat satu jawaban benar. Akan tetapi, bentuk tes pilihan ganda memerlukan pengawasan yang ketat supaya tidak ada kecenderungan peserta didik untuk menyontek, selain itu juga bentuk pilihan ganda cenderung tidak efektif apabila digunakan sebagai alat evaluasi untuk mengukur kemampuan berpikir tingkat tinggi pada peserta didik.

Banyak pendidik mempromosikan penggunaan penilaian alternatif, penilaian untuk pembelajaran yang dapat mengungkapkan proses belajar siswa dengan lebih baik dari penilaian tradisional yang hanya fokus pada hasil belajar siswa (Zessoules & Gardner, 1991; Wiggins, 1998). Mereka percaya bahwa penggunaan penilaian untuk belajar di instruksi kelas dapat memberdayakan siswa sebagai pembelajar dan dengan demikian meningkatkan kinerja siswa (Sadler, 1998; Black & Wiliam, 2004).

Fokus penilaian untuk belajar meningkatkan prestasi siswa (Reeves, 2001) dan siswa belajar daripada mengajar (Harris, 2007). Penilaian untuk pembelajaran juga mencakup umpan balik yang dirancang untuk memberikan informasi langsung, relevan dan bermanfaat bagi siswa dan umpan balik formatif bertujuan untuk menyediakan informasi yang dikomunikasikan kepada siswa untuk mendukung modifikasi pemikiran atau perilaku untuk meningkatkan pembelajaran (Shute, 2008).

Matematika sebagai suatu disiplin ilmu yang secara jelas mengandalkan proses berpikir dipandang sangat baik untuk diajarkan pada anak didik. Di dalamnya terkandung berbagai aspek yang secara substansial menuntun murid untuk berpikir logis menurut pola dan aturan yang telah tersusun secara baku. Sehingga seringkali tujuan utama dari mengajarkan matematika tidak lain untuk membiasakan agar anak didik mampu berpikir logis, kritis dan sistematis. Khususnya berpikir kritis, sangat diperlukan bagi kehidupan

mereka, agar mereka mampu menyaring informasi, memilih layak atau tidaknya suatu kebutuhan, mempertanyakan kebenaran yang terkadang dibaluti kebohongan, dan segala hal yang dapat saja membahayakan kehidupan mereka. Apalagi pada pembelajaran matematika yang dominan mengandalkan kemampuan daya pikir, perlu membina kemampuan berpikir siswa (khususnya berpikir kritis) agar mampu mengatasi permasalahan pembelajaran matematika tersebut yang materinya cenderung bersifat abstrak.

Sekolah justru mendorong siswa memberi jawaban yang benar daripada mendorong mereka memunculkan ide-ide baru atau memikirkan ulang kesimpulan-kesimpulan yang sudah ada. Terlalu sering para guru meminta siswa untuk menceritakan kembali, mendefinisikan, mendeskripsikan, menguraikan, dan mendaftar daripada menganalisis, menarik kesimpulan, menghubungkan, mensintesis, mengkritik, menciptakan, mengevaluasi, memikirkan dan memikirkan ulang. Akibatnya banyak sekolah meluluskan siswa-siswa yang berpikir secara dangkal, hanya berdiri di permukaan persoalan, bukannya siswa-siswa yang mampu berpikir secara mendalam.

Hasil wawancara dengan guru matematika di suatu SMK di Kota Cirebon membuktikan bahwa instrumen penilaian yang digunakan masih mengukur aspek hafalan dan pemahaman. Pencapaian tujuan pembelajaran matematika yang sebenarnya membutuhkan penggunaan instrumen penilaian yang tidak hanya mencakup hafalan dan pemahaman, tetapi juga dibutuhkan penilaian yang melatih keterampilan berpikir (Lissa, 2012).

Instrumen penilaian yang dirancang dengan baik dan sesuai dengan tingkatan kemampuan berpikir dapat meningkatkan daya berpikir siswa, khususnya berpikir kritis. Keterampilan berpikir kritis sangat penting dilatihkan karena keterampilan berpikir ini tidak dibawa sejak lahir (Redhana & Liliarsari, 2008). Pendidikan berpikir di sekolah saat ini khususnya di SMK belum ditangani dengan baik sehingga kecakapan berpikir kritis pada lulusan SMK masih relatif rendah. Rendahnya keterampilan berpikir kritis dan kreatif lulusan pada sekolah dasar sampai dengan perguruan tinggi di Indonesia masih sering dikeluhkan (Reta, 2012). Berdasarkan hal tersebut, dilakukan penelitian pengembangan instrumen asesmen otentik keterampilan berpikir kritis siswa pada pembelajaran matematika.

Instrumen penilaian yang dikembangkan dalam penelitian ini adalah instrumen penilaian yang dapat mengukur keterampilan berpikir kritis siswa. Instrumen penilaian ini didesain untuk meningkatkan keterampilan berpikir siswa. Keterampilan berpikir siswa

dapat dilihat dari jenjang instrumen penilaian yang diujikan dan proporsi ketuntasan. Selain itu, instrumen penilaian keterampilan berpikir kritis pada pembelajaran matematika yang disajikan mengangkat fenomena yang terjadi dalam kehidupan sehari-hari.

Tujuan penelitian adalah untuk 1) menghasilkan instrumen penilaian otentik yang layak digunakan untuk mengukur keterampilan berpikir kritis peserta didik, 2) mengetahui kualitas, kepraktisan instrumen asesmen otentik keterampilan berpikir kritis peserta didik.

METODE PENELITIAN

Penelitian dilakukan di suatu SMK Wahidin, Kota Cirebon, Jawa Barat. Jenis penelitian termasuk *Research and Development* (R&D) yaitu penelitian pengembangan instrumen asesmen otentik keterampilan berpikir kritis. Jenis penelitian R&D yang digunakan dalam penelitian ini mengacu pada Sugiyono (2010) yang diadaptasi sesuai dengan kebutuhan penelitian. Waktu penelitian dimulai dari bulan Juli 2018 sampai bulan Desember 2018.

Tahap penelitian ini dibagi menjadi empat tahap, yaitu tahap pendefinisian, penyusunan desain, pengembangan, dan implementasi. Pendefinisian meliputi tahapan studi literatur dilakukan dengan mencari referensi mengenai kriteria pengembangan keterampilan berpikir kritis serta indikator-indikator keterampilan berpikir kritis.

Desain produk diawali dengan menyusun kisi-kisi soal, menyusun soal, menyusun kunci jawaban, dan validasi desain oleh pakar penelitian pendidikan, pakar keterampilan berpikir kritis, pakar matematika, dan praktisi lapangan. Setelah divalidasi, instrumen penilaian mengalami beberapa kali revisi untuk memperbaiki instrumen penilaian yang dikembangkan sehingga layak untuk diujicobakan di kelas uji coba. Perbaikan dan penyempurnaan instrumen penilaian dilakukan dengan arahan, bimbingan serta masukan dari validator.

Tahap pengembangan dilakukan dengan uji kualitas instrumen yakni dengan menguji validitas dan reliabilitas soal di SMK Wahidin Kota Cirebon. Instrumen dinyatakan valid oleh pakar dan memiliki koefisien reliabilitas dan validitas dengan kategori cukup sampai dengan kategori tinggi, kemudian diujicobakan pada skala terbatas yang melibatkan XI RPL SMK tersebut. Hasil uji coba skala terbatas kemudian dianalisis untuk mengetahui validitas isi menggunakan *Rasch model* digunakan untuk mengetahui kesesuaian item tes (*goodness of fit*) dengan model *Rasch*, reliabilitas soal tes dengan rumus *alphacronbach* menggunakan *software Winsteps* versi 3.73 untuk mendapatkan instrumen penilaian yang lebih reliabel yang kemudian diujicobakan pada skala besar di

kelas XI AK I. Hasil analisis uji coba skala besar didapatkan instrumen penilaian final, kemudian diimplementasikan di kelas XI AK II. Semua sampel diambil secara *purposive sampling*. Tahap pendefinisian, diperoleh data yang meliputi jenis dan kualitas instrumen penilaian yang digunakan di sekolah, kondisi sekolah dan proses pembelajaran matematika. Pada tahap pengembangan, data yang terkumpul adalah hasil validitas dan realibilitas instrumen sedangkan pada tahap implementasi, data yang terkumpul adalah hasil kepraktisan instrumen penilaian. Data tersebut dikumpulkan dengan menggunakan instrumen penelitian yang berupa lembar validasi pakar, lembar angket, lembar *checklist*, lembar aktivitas siswa, tes *essay* analisis, dan tes *problem solving*. Data kualitatif diolah dengan menggunakan teknik penjumlahan sederhana kemudian dilakukan kategorisasi.

HASIL DAN PEMBAHASAN

Penelitian ini merupakan penelitian pengembangan instrumen penilaian otentik. Instrumen penilaian otentik yang dikembangkan yaitu tes tertulis berupa soal uraian untuk mengukur keterampilan proses sains aspek kognitif kategori HOTS dan soal uraian untuk mengukur keterampilan berpikir kritis. Hasil penelitian berdasarkan tahapan-tahapan yang dilakukan selebihnya dijelaskan sebagai berikut:

Pada tahap pendefinisian didapatkan data tentang jenis instrumen asesmen matematika tepatnya materi geometri di sekolah, selain itu juga mengukur aspek hafalan dan pemahaman konsep. Berdasarkan taksonomi kognitif *Bloom* berada pada ranah C1 (hafalan) dan C2 (pemahaman). Kondisi seperti ini tentu tidak lebih baik untuk melatih keterampilan berpikir kritis siswa. Jenis soal dengan tingkat taksonomi *Bloom* yang rendah tidak melatih keterampilan berpikir siswa (Pursitasari & Permanasari, 2012; Ennis, 1993).

Ennis (1993) telah mengembangkan tes uraian untuk mengukur kemampuan berpikir kritis, tetapi tes uraian tersebut bersifat umum. Selain itu, Kartimi & Liliyasi (2012) juga telah mengembangkan tes kemampuan berpikir kritis tetapi pada materi termokimia, Amalia & Susilaningsih (2014) pada materi asam basa serta Amarila, dkk., (2014) dan Jazuli & Wardani (2015) pada mata pelajaran IPA tingkat SMP. Tes berpikir kritis yang berkaitan dengan materi geometri untuk siswa SMK masih belum ada. Oleh karena itu dilakukan pengembangan tes geometri untuk siswa SMK.

Salah satu langkah pada tahap penyusunan desain adalah pemilihan patokan dalam pengembangan tes kemampuan berpikir kritis. Patokan yang dipilih adalah lima aspek kemampuan berpikir kritis yang dikemukakan oleh Ennis & Weir (1985), yaitu: (a)

memberikan penjelasan dasar, (b) membangun keterampilan dasar, (c) menyimpulkan, (d) memberikan penjelasan lanjut, dan (e) strategi dan taktik. Kelima aspek tersebut kemudian menjadi patokan dalam mengembangkan butir soal. Dari kelima aspek kemampuan berpikir kritis yang menjadi patokan. Terdapat sepuluh butir soal uraian yang dikembangkan dalam penelitian ini. Sepuluh soal tersebut mewakili lima aspek kemampuan berpikir kritis. Kelima aspek kemampuan berpikir kritis dan rincian butir soal yang mewakilinya ditunjukkan pada tabel 1. Tahap berikutnya adalah pengembangan. Dalam tahap ini ada validasi pakar. Validasi yang dilakukan adalah validitas isi dari instrumen penilaian. Instrumen keterampilan berpikir hendaknya memiliki validitas konstruk yang baik sebelum digunakan (Ennis & Weir, 1985; Docktor & Heller, 2009). Oleh karena itu, validasi pakar menjadi bagian yang penting untuk memulai pengembangan.

Validasi konstruk meliputi empat aspek, yaitu: (a) kesesuaian butir soal dengan indikator, (b) tingkat kesukaran butir soal (konsep soal), (c) penggunaan bahasa dalam butir soal, dan (d) kebenaran konsep kunci jawaban. Selain itu, validasi isi juga dilakukan untuk mendapatkan saran terhadap butir soal tes dari ahli. Analisis data dari hasil validasi isi dilakukan dengan metode deskripsi rata-rata. Selain itu, butir soal tes juga direvisi berdasarkan saran dari ahli. Hasil validasi konstruk oleh *expert judgement* digunakan untuk memperbaiki instrumen penilaian otentik keterampilan proses sains dan instrumen penilaian otentik keterampilan berpikir kritis. Perbaikan dilakukan sesuai dengan saran validator.

Validitas dinyatakan baik dengan kategori koefisien validitas berkisar antara valid sampai dengan sangat valid. Hasil validitas konstruk *expert judgment* dapat dilihat pada tabel 2.

Tabel 1. Lima Aspek Kemampuan Berpikir Kritis dan Rincian Butir Soal yang Mewakilinya

| No | Aspek Kemampuan Berpikir Kritis | Butir Soal |
|----|---------------------------------|------------|
| 1 | Memberikan penjelasan dasar | 1, 2 |
| 2 | Membangun keterampilan dasar | 3, 4 |
| 3 | Menyimpulkan | 5, 6 |
| 4 | Memberikan penjelasan lanjut | 7, 8, 9 |
| 5 | Strategi dan taktik | 10 |

Tabel 2. Rangkuman Hasil Validasi Instrumen Asesmen Unjuk Kerja Pada Pembelajaran Matematika Sekolah Dasar Kota Cirebon

| No. | Aspek Yang Dinilai | Kriteria Penilaian | Validator Ahli | | | | Skor 20 | Kualitas |
|-------------|--|--------------------|----------------|----|----|----|---------|----------|
| | | | 1 | 2 | 3 | 4 | | |
| 1 | Kesesuaian butir soal dengan indicator | 1 | 4 | 5 | 5 | 4 | 18 | SB |
| | | 2 | 5 | 4 | 4 | 4 | 17 | |
| | | 3 | 4 | 4 | 5 | 4 | 17 | |
| 2 | Konsep Soal | 4 | 4 | 5 | 4 | 4 | 17 | SB |
| | | 5 | 5 | 5 | 5 | 5 | 20 | |
| | | 6 | 5 | 4 | 4 | 5 | 19 | |
| 3 | Bahasa | 7 | 4 | 5 | 4 | 5 | 18 | SB |
| | | 8 | 4 | 4 | 3 | 4 | 15 | |
| 4 | Kebenaran Kunci jawaban | 9 | 3 | 4 | 3 | 4 | 14 | SB |
| | | 10 | 4 | 4 | 4 | 5 | 17 | |
| Jumlah Skor | | | 42 | 43 | 41 | 45 | 172 | SB |

Selain nilai persentase keidealan hasil validasi umum dari ke-empat penilai terhadap instrumen asesmen berpikir kritis yaitu dapat dilihat dari nilai *expert judgment*, yang mendapatkan skor maksimal 20 ada pada aspek konsep soal. Selain data yang dipaparkan pada tabel juga diperoleh data tertulis yang berasal dari kolom catatan dan data verbal yang ditranskripsikan dari hasil wawancara dengan ahli dan praktisi tentang Beberapa hal pokok yang menjadi masukan dari ke-empat penilai tersebut diantaranya adalah: (1) Kalimat pada instrumen hendaknya langsung ke akar permasalahan, tidak bertele-tele; (2) kebenaran kunci jawaban masih kurang benar; (3) Instrumen Hendaknya mengukur kompetensi spesifik yang diharapkan muncul dalam pembelajaran.

Validasi isi tersebut hanya terbatas pada kesesuaian materi dengan kemampuan berpikir kritis. Hasil validasi isi belum bisa menunjukkan bagaimana respon msiswa terhadap butir soal tersebut. Oleh karena itu, perlu adanya validasi lanjutan untuk mengetahui tingkat validitas dan reliabilitas butir soal yang telah dikembangkan serta mengetahui respon siswa terhadap butir soal.

Validasi empiris dilakukan dengan subjek siswa SMMK Wahidin Cirebon kelas XI AK I sebanyak 30 siswa. Siswa diberi waktu 100 menit untuk mengerjakan butir soal tes secara mandiri. Setelah siswa mengerjakan tes, jawaban siswa kemudian dikoreksi dan dianalisis menggunakan *software Winsteps versi 3.73*. Hasil analisisnya ditunjukkan pada Gambar 1.

| SUMMARY OF 30 MEASURED PERSON | | | | | | | | |
|--|-------------|---------|---------|-------------|-------|--------------------|--------|------|
| | TOTAL SCORE | COUNT | MEASURE | MODEL ERROR | INFIT | | OUTFIT | |
| | | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| MEAN | 60.0 | 10.0 | .30 | .28 | .99 | .0 | .99 | .0 |
| S.D. | 6.1 | .0 | .46 | .01 | .36 | .9 | .35 | .9 |
| MAX. | 70.0 | 10.0 | 1.09 | .30 | 1.95 | 1.8 | 1.90 | 1.8 |
| MIN. | 44.0 | 10.0 | -.87 | .27 | .27 | -2.2 | .28 | -2.2 |
| REAL RMSE | .30 | TRUE SD | .36 | SEPARATION | 1.20 | PERSON RELIABILITY | .81 | |
| MODEL RMSE | .28 | TRUE SD | .37 | SEPARATION | 1.34 | PERSON RELIABILITY | .79 | |
| S.E. OF PERSON MEAN = .09 | | | | | | | | |
| PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00 | | | | | | | | |
| CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .82 | | | | | | | | |
| SUMMARY OF 10 MEASURED ITEM | | | | | | | | |
| | TOTAL SCORE | COUNT | MEASURE | MODEL ERROR | INFIT | | OUTFIT | |
| | | | | | MNSQ | ZSTD | MNSQ | ZSTD |
| MEAN | 180.0 | 30.0 | .00 | .16 | .99 | .0 | .99 | .0 |
| S.D. | 22.3 | .0 | .54 | .00 | .20 | .8 | .20 | .8 |
| MAX. | 202.0 | 30.0 | 1.08 | .17 | 1.28 | 1.1 | 1.28 | 1.1 |
| MIN. | 135.0 | 30.0 | -.56 | .15 | .66 | -1.4 | .68 | -1.3 |
| REAL RMSE | .17 | TRUE SD | .52 | SEPARATION | 3.09 | ITEM RELIABILITY | .91 | |
| MODEL RMSE | .16 | TRUE SD | .52 | SEPARATION | 3.23 | ITEM RELIABILITY | .91 | |
| S.E. OF ITEM MEAN = .18 | | | | | | | | |
| UMEAN=.0000 USCALE=1.0000 | | | | | | | | |
| ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00 | | | | | | | | |
| 300 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 920.84 with 255 d.f. p=.0000 | | | | | | | | |
| Global Root-Mean-Square Residual (excluding extreme scores): 1.1377 | | | | | | | | |

Gambar 1. *Person Reliability*

Informasi penting yang dapat diperoleh dari tabel *Summary Statistics* adalah *Person Reliability* dan *Item Reliability*, nilai alpha Cronbach dan *Person Measure*. Dapat diketahui besarnya *Person Reliability* pada keterampilan berpikir kritis siswa Matematika sebesar 0,81 sedangkan *Item Reliability* sebesar 0,91. Besarnya *Alpha Cronbach* yaitu 0,82. *Person Measure* 0,16 logit menunjukkan rata-rata kemampuan berpikir kritis siswa di atas item. Atau dengan kata lain kecenderungan siswa lebih banyak menjawab benar dari semua soal yang dikerjakan. Nilai *Mean IMNSQ person* dan *item* adalah 0,99 sedangkan *mean OMNSQ person* dan *item* sebesar 0,99, nilai keduanya kurang dari 1,0 yang artinya menunjukan pengamatan yang terlalu mudah ditebak (overfit model), (Fisher, 2007).

Selanjutnya untuk hasil analisis butir soal didapat dengan menjumlahkan nilai *mean* dan S.D. *Infit mean square (MNSQ)* yang hasilnya adalah 1,19. Nilai logit per item bila lebih besar dari 1, 19 maka butir tersebut dinyatakan dalam kategori tidak fit dengan model *Rasch*.

Dari gambar 2, yaitu nilai pada kolom *OUTFIT MNSQ* adalah berada pada rentang 0,5 sampai dengan 1,5 sehingga dapat disimpulkan bahwa 10 item adalah fit terhadap model *Rasch*. Untuk memeriksa item yang tidak sesuai (*outliers* atau *misfits*) dengan meninjau hal-hal berikut ini (Azrilah, 2010).

1. Nilai Outfit Mean Square (MNSQ) yang diterima berkisar 0,5 – 1,5
2. Nilai Outfit Z-Standard (ZSTD) yang diterima berkisar -2,0 – 2,0
3. Nilai Point Measure Correlation (Pt Mean Corr) berkisar 0,4 sampai 0,85

Dari hasil analisis tersebut dapat diketahui bahwa semua item fit, sehingga soal tidak perlu direvisi atau dieliminasi.

Oleh karena itu bisa disimpulkan instrumen asesmen otentik berpikir kritis siswa valid dan reliabel dilihat dari validitas konstruk dan validitas isi instrumen.

ITEM STATISTICS: MISFIT ORDER

| ENTRY NUMBER | TOTAL SCORE | TOTAL COUNT | MEASURE | MODEL | | INFIT | | OUTFIT | | PT-MEASURE | | EXACT MATCH | | ITEM |
|--------------|-------------|-------------|---------|-------|------|-------|------|--------|-------|------------|------|-------------|------|--------|
| | | | | S. E. | MNSQ | ZSTD | MNSQ | ZSTD | CORR. | EXP. | OBS% | EXP% | | |
| 9 | 202 | 30 | -.56 | .17 | 1.28 | 1.1 | 1.18 | 1.1 | A | .42 | .46 | 36.7 | 35.5 | Item9 |
| 5 | 184 | 30 | -.08 | .16 | 1.27 | 1.1 | 1.17 | 1.1 | B | .51 | .47 | 23.3 | 33.8 | Item5 |
| 1 | 196 | 30 | -.40 | .16 | 1.13 | .6 | 1.12 | .6 | C | .43 | .46 | 36.7 | 34.9 | Item1 |
| 10 | 139 | 30 | .99 | .15 | 1.07 | .4 | 1.07 | .4 | D | .45 | .48 | 36.7 | 31.7 | Item10 |
| 8 | 189 | 30 | -.21 | .16 | 1.04 | .3 | 1.03 | .2 | E | .40 | .47 | 36.7 | 34.2 | Item8 |
| 4 | 181 | 30 | -.01 | .16 | 1.01 | .1 | 1.00 | .1 | e | .38 | .48 | 26.7 | 33.4 | Item4 |
| 2 | 135 | 30 | 1.08 | .15 | .97 | -.1 | .96 | -.1 | d | .54 | .48 | 30.0 | 31.7 | Item2 |
| 6 | 194 | 30 | -.34 | .16 | .85 | -.5 | .82 | -.7 | c | .58 | .47 | 23.3 | 34.7 | Item6 |
| 3 | 195 | 30 | -.37 | .16 | .66 | -1.4 | .69 | -1.3 | b | .62 | .46 | 43.3 | 35.0 | Item3 |
| 7 | 185 | 30 | -.11 | .16 | .67 | -1.4 | .68 | -1.3 | a | .43 | .47 | 33.3 | 33.8 | Item7 |
| MEAN | 180.0 | 30.0 | .00 | .16 | .99 | .0 | .99 | .0 | | | | 32.7 | 33.9 | |
| S. D. | 22.3 | .0 | .54 | .00 | .20 | .8 | .20 | .8 | | | | 6.3 | 1.2 | |

Gambar 2. MISFIT Order

Tahap Selanjutnya adalah Implementasi. Pada tahap ini memaprkan kepraktisan instrumen asesmen otentik berpikir kritis siswa.Kepraktisan suatu tes merupakan indikator kualitas suatu alat ukur tergolong baik atau tidak. Hasil analisa kepraktisan instrumen asesmen otentik berpikir kritis siswa berdasarkan data yang diperoleh dengan memberikan angket kuesioner kepada keempat penilai (guru) yang mengujicobakan penggunaan instrumen. Secara empirik keempat penilai diminta mengisi koesioner dengan lima jenis pertanyaan dalam bentuk rubrik dengan rentang nilai 1 sampai 5, yang berhubungan dengan kepraktisan penggunaan. Data hasil jawaban responden selanjutnya dianalisis secara statistik dengan menggunakan formula T Skor. Berdasarkan anáalisis data dapat diinterpretasikan bahwa melalui uji skor baku (stándar) dengan skor T terhadap kepraktisan penggunaan instrumen penilaian unjuk pada pembelajaran matematika dilihat pada tabel 3.

Tabel 3. Data Penilaian Rater Terhadap KepraktisanInstrumen Asesmen Otentik Berpikir Kritis Siswa

| Aspek Penilaian | Kriteria n Penilaian n | Penilai | | | Σ Skor | Σ Per Aspek | Rata-Rata | Persentase Keidealn | Kategori Kualitas |
|-----------------|------------------------|---------|----|-----|--------|-------------|-----------|---------------------|-------------------|
| | | I | II | III | | | | | |
| Objektivitas | 1 | 4 | 5 | 4 | 13 | 39 | 13 | 95,53% | Sangat Baik |
| | 2 | 4 | 4 | 4 | 12 | | | | |
| | 3 | 5 | 5 | 4 | 14 | | | | |
| Kesistematican | 4 | 3 | 3 | 5 | 11 | 24 | 11,5 | 83% | Sangat Baik |
| | 5 | 4 | 4 | 4 | 12 | | | | |
| Konstruksi | 6 | 5 | 4 | 5 | 14 | 27 | 13,5 | 93% | Sangat |

| | | | | | | | | | |
|-----------------------|----|----|----|----|-----|------|------------|--------|-------------|
| | 7 | 4 | 4 | 5 | 13 | | | | Baik |
| Kebahasaan | 8 | 5 | 4 | 4 | 13 | 13 | 13 | 86,7% | Sangat Baik |
| | 9 | 4 | 4 | 5 | 13 | | | | |
| Kepraktisan | 10 | 5 | 5 | 4 | 14 | 41 | 13,6666667 | 91,13% | Sangat Baik |
| | 11 | 4 | 5 | 5 | 14 | | | | |
| Jumlah skor | | 47 | 47 | 49 | 143 | 144 | 64,6666667 | 90,51% | Sangat Baik |
| Jumlah skor Rata-Rata | | | | | | 67,5 | | | Baik |

Terlihat dari Tabel 3 secara umum guru-guru menilai instrumen asesmen otentik berpikir kritis siswa memiliki subyektivitas, kesistematian, konstruksi, kebahasaan dan kepraktisan yang baik. Hal ini tergambar dari skor T dari masing *rater* I sebesar 47, *rater* II diperoleh skor T sebesar 47, dan dari *rater* III diperoleh skor T sebesar 49. Dengan demikian menurut kriteria kepraktisan *formulaGlicman* (Arikunto, 2010) instrumen asesmen otentik berpikir kritis siswa ini dapat dikatakan secara umum dinilai praktis oleh para guru (*rater*) dalam menilai tingkat kualitas kepraktisan instrumen asesmen otentik berpikir kritis siswa SMK pada pembelajaran matematika materi geometri.

KESIMPULAN

Penelitian ini menghasilkan Instrumen penilaian otentik yang layak digunakan untuk mengukur keterampilan berpikir kritis peserta didik kelas XI SMK Wahidin Kota Cirebon tahun ajaran 2017/2018 pada materi geometri. Kelayakan instrumen penilaian otentik ini didasarkan pada hasil analisis sebagai berikut: (a) instrumen telah memenuhi syarat validitas isi oleh *expert judgement* dengan kategori sangat baik dan sebanyak 10 butir soal keterampilan berpikir kritis telah mendapatkan bukti empiris *fit* dengan *RaschModel* berdasarkan tiga parameter yaitu *OUTFIT MNSQ*, *ZSTD*, dan *PtMean Corr*; (b) berdasarkan *index sparasi item*, instrumen penilaian otentik berpikir kritis yang dikembangkan tergolong reliabel. Kualitas implementasi asesmen otentik keterampilan berpikir kritis pada pembelajaran matematika adalah praktis.

DAFTAR PUSTAKA

- Amalia, N. F., & Susilaningsih, E. (2014). Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Siswa SMA pada Materi Asam Basa. *Jurnal Inovasi Pendidikan Kimia*, Vol.8, No.2, 1380-1389.
- Amarila, R. S., Habibah, N. A., & Widiyatmoko, A. (2014). Pengembangan Alat Evaluasi Kemampuan Berpikir Kritis Siswa pada Pembelajaran IPA Terpadu Model Webbed Tema Lingkungan. *Unnes Science Education Journal*, 3(2), 563-569.
- Arikunto, S. (2010). *Dasar-dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara.

- Azrilah, A. A. (2010). *Rasch Model Fundamentals: Scale Construct and Measurement Structure*. Perpustakaan Negara Malaysia, Kuala Lumpur: Integrated Advance Planning Sdn. Bhd.
- Black, P., & Wiliam, D. (2004). *The Formative Purpose: Assessment Must First Promote Learning*. In M. Wilson (Ed.), *Towards Coherence Between Classroom Assessment and Accountability*. Chicago: University of Chicago Press.
- Docktor, J., & Heller, K. (2009). *Robust Assessment Instrument for Student Problem Solving*. In Proceedings of the NARST 2009 Annual Meeting, Garden Grove, CA, 1-19.
- Ennis, R. H., & Weir, E. (1985). *The Ennis Weir Critical Thinking Essay Test*. Pacific Grove, CA: Midwest Publication.
- Ennis, R. H. (1993). Critical Thinking Assessment. *Journal College of Education*, Vol.32, No.3, 179-186.
- Fisher, W. P. (2007). Rating Scale Instrument Quality Criteria. *Rasch Measurement Transaction*, Vol. 21, No.1, 1087-1096.
- Harris, L. (2007). Employing Formative Assessment in the Classroom. *Improving Schools*, 10(3), 249-260.
- Jazuli, M & Wardani, S. (2015). Pengembangan Alat Evaluasi IPA Terpadu Topik Perubahan Materi Berbasis Kontekstual untuk Mengukur Kemampuan Berpikir Kritis Siswa. *Unnes Science Education Journal*, Vol.4, No.2, 912-918.
- Johnson, B. (2007). *Contextual Teaching & Learning*. Bandung: Mizan Media Utama.
- Kartimi & Liliyasi. (2012). Pengembangan Alat Ukur Berpikir Kritis pada Konsep Termokimia untuk Siswa SMA Peringkat Atas dan Menengah. *Jurnal Pendidikan IPA Indonesia*, Vol.1, No.2, 21-26.
- Lissa. (2012). Pengembangan Instrumen Penilaian Keterampilan Berpikir Tingkat Tinggi Materi Sistem Respirasi Dan Ekskresi. *Jurnal Lembaran Ilmu Kependidikan*, Vol.41, No.1, 27-32.
- Pursitasari, I. D., & Permanasari. A. (2012). Analisis Pemahaman Konsep dan Kesulitan Mahasiswa untuk Pengembangan Program Perkuliahan Dasar-Dasar Kimia Analitik Berbasis *Problem Solving*. *Jurnal Pendidikan IPA Indonesia*, Vol.1, No.1, 98-101.
- Redhana, I. W., & Liliyasi. (2008). Program Pembelajaran Keterampilan Berpikir Kritis pada Topik Laju Reaksi untuk Siswa SMA. *Forum Kependidikan*, 27(2), 103-112.
- Reeves, D. B. (2001). Standards Make a Difference: The Influence of Standards in Classroom Assessment. *NASSP Bulletin*, 85(5), 5-12.
- Reta, I. K. (2012). Pengaruh Model Pembelajaran Berbasis Masalah terhadap Keterampilan Berpikir Kritis Ditinjau dari Gaya Kognitif Siswa. *Jurnal Pendidikan*, Vol.26, No.1, 1-16.
- Sadler, R. (1998). Formative Assessment: Revisiting the Territory. *Assessment in Education Journal: Principles, Policy and Practice*, 5(1), 77-85.
- Sugiyono. (2010). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(2), 153-189.
- Wiggins, G. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco, CA: Jossey-Bass
- Zessoules, R & Gardner, H. (1991). *Authentic Assessment: Beyond the Buzzword and Into The Classroom*, In V, Perrone (Ed.), *Expanding Student Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.