



Analisis Iteaman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan

Sri Nuryanti *, Muhammad Masykuri, E. Susilowati

Universitas Sebelas Maret. Jalan Ir. Sutami No.36 A, Kota Surakarta, 57126, Indonesia

* Corresponding Author. Email: eyantie@gmail.com

Received: 2 October 2018; Revised: 28 November 2018; Accepted: 17 December 2018

Abstrak

Tujuan penelitian ini adalah untuk menganalisis instrumen kemampuan berpikir kritis. Penelitian ini merupakan penelitian kuantitatif. Data diperoleh dengan mengembangkan instrumen penilaian kemampuan berpikir kritis. Responden sebanyak 90 peserta didik yang diambil dari 3 SMK yang ada di kota Surakarta. Analisis data menggunakan perangkat lunak (*software*) analisis *Iteaman versi 3.0* dan *Winstep versi 3.73*. Metode yang digunakan dalam analisis ini yaitu metode tes klasik yaitu *Iteaman* dan pemodelan *Rasch*. Hasil analisis dengan pemodelan *Iteaman* diperoleh nilai *Alpha Cronbach* 0,80 berarti reliabilitas dari soal ini tinggi. Sedangkan dengan model *Rasch Alpha Cronbach* 0,74 untuk reliabilitas dan 0,95 untuk reliabilitas *item* yang berarti dapat diandalkan untuk mengukur ketrampilan berpikir kritis. Validitas *item* 0,93 menunjukkan soal dapat mengukur kemampuan berpikir kritis hal ini diperkuat oleh nilai *separation* 4,34. Nilai *separation* menunjukkan *item* soal ini mempunyai sebaran respon yang baik. Instrumen kemampuan berpikir kritis dapat digunakan untuk penelitian selanjutnya.

Kata Kunci: berpikir kritis, instrumen, iteaman, model Rasch

Iteaman and rasch model analysis on the development of critical thinking instruments for vocational school students

Abstract

The aimed of this study is to analyze the instrument of critical thinking skills. This research is quantitative research. Data was obtained by developing instruments for assessing critical thinking skills. Respondents were 90 students that taken from 3 Vocational Schools in Surakarta. Data analysis was using Iteaman analysis software version 3.0 and Winstep version 3.73. The method used in this analysis were the classic test method, iteaman and rasch models. The results of the analysis with iteaman modeling obtained the Alpha Cronbach value of 0,800 means that the reliability of this problem was high. Whereas the Rasch Alpha Cronbach model is 0.74 for person reliability and r0.95 for item reliability which means it can be relied upon to measure critical thinking skills. The validity of the item 0.93 shows the instrument can measure critical thinking skills and this is reinforced by the value of separation 4.34. The value of separation showed this item has a good distribution of responses. The instrument of critical thinking skills can be used for further research.

Keywords: *critical thinking, instruments, iteaman, rasch model*

How to Cite: Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteaman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224-233. doi:<https://doi.org/10.21831/jipi.v4i2.21442>

 <https://doi.org/10.21831/jipi.v4i2.21442>

PENDAHULUAN

Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, Pasal 1 ayat 1 menyatakan bahwa pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan po-

tensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara. Dari pernyataan tersebut menunjukkan bahwa pendidikan tidak hanya untuk meningkatkan pengetahuan peserta saja, melainkan juga

mempersiapkan ketrampilan atau kemampuan yang diperlukan dalam kehidupan mendatang.

Sekolah Menengah Kejuruan (SMK) merupakan jenjang pendidikan yang ada di Indonesia yang bertujuan untuk menghasilkan tenaga kerja madya sesuai dengan keahlian yang ditekuni dan mampu mengikuti perkembangan informasi dan teknologi. Hal ini sejalan dengan penjelasan pasal 15 Undang-Undang Sistem Pendidikan Nasional tahun 2003. SMK merupakan pendidikan menengah yang mempersiapkan peserta didik terutama bekerja pada bidang tertentu. Dalam sebuah artikel yang dimuat di Kompas bulan Agustus 2017 mengatakan belum semua SMK mampu memenuhi tuntutan dunia kerja sesuai dengan spesifikasinya. Jumlah pengangguran dari lulusan pendidikan vokasi, baik sekolah menengah kejuruan maupun diploma I-III, membengkak dalam setahun terakhir. Sekitar 11,41 persen pengangguran terbuka pada Agustus 2017 memiliki tingkat pendidikan menengah kejuruan (Kompas, 2017). Dari artikel tersebut dapat diprediksi bahwa kemampuan lulusan SMK masih kurang untuk memasuki dunia industri.

Saat ini dunia industri memasuki era revolusi industri generasi keempat. *World Economic Forum* pada Januari 2016 memperkirakan ada sekitar 35% keahlian yang dianggap penting saat ini kelak akan berubah. Selanjutnya juga dirangkum dari *World Economic Forum (WEF)*, pada tahun 2020 dimana era revolusi industri generasi keempat dimulai ada sepuluh *softskill* yang harus dimiliki untuk menjawab tantangan dunia industri (Schwab, 2016). *Softskill* tersebut adalah menyelesaikan permasalahan yang kompleks/sulit (*complex problem solving*), berpikir kritis (*critical thinking*), kreatifitas (*creativity*), manajemen SDM (*people management*), koordinasi (*coordinating*), kecerdasan emosional (*emotional intelligence*), pengambilan keputusan (*judgment and decision making*), orientasi pada layanan (*service orientation*), negosiasi (*negotiation*) dan kelenturan berpikir (*cognitive flexibility*).

Revolusi industri 4.0 ditanggapi pemerintah dengan adanya Instruksi Presiden (Inpres) Nomor 9 Tahun 2016 tentang Revitalisasi Sekolah Menengah Kejuruan (SMK). Dunia pendidikan khususnya SMK sangat terbantu karena akan terciptanya sinergi antar instansi dan lembaga terkait sesuai dengan tugas dan fungsi masing-masing dalam usaha mengangkat kualitas SMK. Dalam revitalisasi ini kerjasama dunia industri dengan SMK menjadi

perhatian yang besar, dengan upaya ini diharapkan lulusan SMK dapat mengisi kebutuhan lapangan pekerjaan di industri. Perbaikan dan penyesuaian kurikulum SMK akan memantapkan model kesesuaian dan keterkaitan (*link and match*) dengan dunia Usaha/Industri (Hadam, Rahayu, & Ariyadi, 2018)

Selain berada dalam revolusi industri 4.0 pendidikan saat ini juga dihadapkan pada tuntutan abad 21. Menurut hasil penelitian yang dilakukan oleh lebih dari 250 peneliti dari 60 institusi dunia yang tergabung dalam ATC21S (*Assessment & Teaching of 21st Century Skills*) 4 ketrampilan yang harus dimiliki yaitu; *Communication, Collaboration, Critical Thinking and Problem Solving, Creativity and Innovation*. Kemampuan berpikir kritis merupakan salah satu ketrampilan sangat diperlukan saat peserta didik lulusan SMK memasuki dunia kerja pada abad 21 (Trilling & Fadel, 2009).

Apabila orang mempunyai kemampuan berpikir kritis akan kelihatan dari karakter orang tersebut saat menghadapi masalah. Karakter itu akan kelihatan saat orang tersebut berbicara, bertindak maupun memberikan argumen maupun solusi terhadap suatu masalah. *Facione* mengatakan ada enam kecakapan berpikir kritis utama yang terlibat di dalam proses berpikir kritis. Kecakapan-kecakapan tersebut adalah interpretasi, analisis, evaluasi, inference, penjelasan dan regulasi diri (Facione, 2011a; Peter & Facione, 1989). Untuk mengukur kemampuan berpikir kritis diperlukan suatu instrumen yang mampu mengukur aspek-aspek dalam berpikir kritis. Kemampuan berpikir kritis dapat diukur melalui tes tertulis, lisan dan observasi. Pada penelitian ini kemampuan berpikir kritis tes kemampuan berpikir kritis diadaptasi dan diterjemahkan dari Starkey meliputi aspek menarik kesimpulan, mengenali asumsi, deduksi, interpretasi, analisis, evaluasi atas argumen (Lai, 2011; Starkey, 2009). Kemampuan kritis diperlukan peserta didik untuk menghadapi masalah yang dihadapinya. Kemampuan untuk menganalisis suatu masalah membuat kesimpulan lalu penempatan diri yang tepat sangat diperlukan oleh lulusan SMK untuk menghadapi revolusi industri 4.0.

Dari kebutuhan kompetensi revolusi industri 4.0 dan pendidikan abad 21 tersebut tampak bahwa menyelesaikan masalah dan berpikir kritis merupakan kemampuan yang harus dimiliki seseorang untuk memasuki dunia industri. Berdasarkan hal tersebut maka peneliti akan mengembangkan instrumen yang dapat meng-

ukur kemampuan berpikir kritis. Salah satu indikasi keberhasilan proses belajar adalah skor nilai yang diperoleh peserta didik. Untuk memperoleh nilai tersebut seorang pengajar mengembangkan instrumen. Instrumen disusun berdasarkan kisi-kisi yang diambil dari 6 aspek kemampuan berpikir kritis. Instrumen yang dikembangkan kejadian umum dan bidang teknik yang dihadapi oleh peserta didik. Suatu instrumen yang baik hasilnya dapat dipercaya dan mampu mengukur apa yang seharusnya diukur. Untuk mengetahui kualitas dari instrumen yang disusun maka dilakukan analisis terhadap instrumen tersebut (Zhou, Huang, & Tian, 2013).

Analisis dilakukan untuk mengetahui validitas dan reliabilitas instrumen. Validitas berkenaan dengan ketepatan alat penilaian terhadap apa yang dinilai sehingga betul-betul menilai apa yang seharusnya dinilai hasil penilaian bentuknya skor angka. Terdapat 4 validitas yaitu validitas isi, validitas bangun pengertian (*construct validity*), validitas ramalan dan validitas kesamaan. Validitas isi berkenaan dengan kesanggupan instrumen untuk mengukur isi yang seharusnya, dalam hal ini kemampuan instrumen mengukur kemampuan berpikir kritis peserta didik. Validitas bangun berarti kesanggupan instrumen mengukur pengertian yang terkandung dalam materi yang diukur. Kemampuan meramalkan atau memprediksikan ciri, perilaku atau kriteria tertentu disebut dengan validitas ramalan. Sedangkan validitas kesamaan artinya tes tersebut memiliki kesamaan dengan tes yang sudah dibakukan (Sudjana, 2005)

Pada penelitian ini instrumen akan dianalisis dengan teori klasik menggunakan *software iteman* dan analisis pemodelan *Rasch* dengan menggunakan *software winstep*. Kedua analisis dilakukan instrumen yang dihasilkan memiliki validitas dan reliabilitas yang cukup baik untuk mengukur kemampuan berpikir kritis. Instrumen ini selanjutnya dapat digunakan sesuai dengan kebutuhan penelitian selanjutnya.

METODE

Dalam penelitian ini digunakan 9 langkah penelitian dan pengembangan (*R&D*) menurut Borg dan Gall (2007). Instrumen ini terlebih dahulu divalidasi oleh 8 ahli untuk menentukan validitas konstruk dan validitas isinya (Morganna, Pinto, Federal, & Grande, 2007). Data diperoleh dengan mengembangkan instrumen penilaian kemampuan berpikir kritis. Responden sebanyak 90 peserta didik yang diambil dari 3 SMK yang ada di kota Surakarta. Analisis

instrumen kemampuan berpikir kritis menggunakan perangkat lunak (*software*) analisis *Iteman* versi 3.0 dan *Winstep* versi 3.73.

Instrumen yang digunakan adalah soal kemampuan berpikir kritis yang berisi 20 soal pilihan ganda. Dua puluh soal ini dikembangkan dari enam aspek kemampuan berpikir kritis yang disampaikan oleh Facione (Facione, 2011). Enam aspek tersebut meliputi interpretasi, analisis, evaluasi, *interference*, penjelasan dan regulasi diri. Instrumen kemampuan berpikir dikembangkan dari pengetahuan umum dan pengetahuan teknik. Responden menentukan jawaban yang benar. Sementara kunci jawaban disediakan.

Secara isi instrumen ini divalidasi ahli media, ahli materi, praktisi pendidikan (guru) dan angket teman sejawat. Pada penelitian ini validitas isi ditentukan dengan menggunakan formula *Aiken* sebagai berikut (Azwar, 2015, p 113)

$$V = \frac{S}{[n(c - 1)]}$$

Keterangan:

V = validitas

S = r - lo

r = angka yang diberikan oleh seorang penilai

lo = angka penilaian validitas yang terendah

c = angka penilaian validitas yang tertinggi

n = banyaknya penilai validitas dari Aiken

Dari analisis ini dapat diketahui item soal yang baik (valid), perlu direvisi atau item soal yang harus dibuang.

Sampel penelitian ini diambil dari tiga SMK yang ada di kota Surakarta, satu SMK kriteria tinggi, satu SMK kriteria sedang dan satu SMK kriteria rendah. Penilaian kriteria SMK ini berdasarkan data nilai skor Ujian Nasional SMK tahun pelajaran 2016/2017. Sampel dari 3 sekolah ini diambil secara acak. Penelitian ini dibantu oleh guru pengampu dari masing-masing sekolah.

Iteman merupakan program komputer yang digunakan untuk menganalisis butir soal secara klasik. Program ini termasuk satu paket program dalam MicroCATn yang dikembangkan oleh Assessment Systems Corporation pada tahun 1982. Program ini dapat digunakan untuk: (1) menganalisis data file (format ASCII) jawaban butir soal yang dihasilkan melalui manual entry data atau dari mesin scanner; (2) menskor dan menganalisis data soal pilihan ganda dan skala Likert untuk 30.000 peserta didik dan 250 butir soal; (3) menganalisis sebuah tes yang

terdiri dari 10 skala (subtes) dan memberikan informasi tentang validitas setiap butir (daya pembeda, tingkat kesukaran, proporsi jawaban pada setiap option), reliabilitas (KR-20/Alpha), standar error of measurement, mean, variance, standar deviasi, skew, kurtosis untuk jumlah skor pada jawaban benar, skor minimum dan maksimum, skor median, dan frekuensi distribusi nilai skor.

Prosedur analisis item soal berdasarkan teori tes klasik yaitu reliabilitas, tingkat kesukaran dan daya beda soal. Uji reliabilitas menggunakan rumus *Alpha Cronbach*, klasifikasi reliabilitas: 0,00-0,20 (sangat rendah); 0,21-0,40 (rendah); 0,41-0,70 (sedang); 0,71-0,90 (tinggi); 0,91-1,00 (sangat tinggi) (Arikunto, 2016).

Daya pembeda soal merupakan kemampuan sebuah soal untuk membedakan antara peserta didik yang berkemampuan tinggi dan berkemampuan rendah. Indeks daya pembeda soal yang tinggi menunjukkan semakin mampu soal untuk membedakan kemampuan peserta didik. Jika daya pembeda soal bernilai negatif berarti lebih banyak kelompok peserta didik yang belum memahami materi tersebut Kualifikasi sebagai berikut: 0,00-0,20 (jelek); 0,21-0,40 (cukup); 0,41-0,70 (baik); 0,71-1,00 (baik sekali); negatif (jelek) (Arikunto, 2016).

Butir-butir item soal yang baik apabila butir-butir item tersebut tidak terlalu sukar dan tidak pula terlalu mudah dengan kata lain derajat kesukaran item itu adalah sedang atau cukup. Angka indeks kesukaran dapat diperoleh dengan menggunakan rumus yang dikemukakan oleh Robert L. Thorndike. Klasifikasi tingkat kesukaran soal disajikan sebagai berikut: 0,00 – 0,30 (sulit); 0,31-0,70 (sedang); 0,71-1,00 (mudah) (Ngalimun, 2018.).

Analisis model Rasch dikemukakan oleh Georg Rasch tahun 1960-an dan dipopulerkan oleh Ben Wirght. Data mentah yang digunakan merupakan data dikotomi yang menunjukkan kemampuan peserta didik. Dalam analisis model Rasch ini dalam satu model dapat diketahui hubungan antara soal dan peserta didik (Sumintono & Widhiarso, 2015).

Data yang diperoleh diolah dengan menggunakan *software Winsteps*. Pengolahan dengan menggunakan model Rasch ini akan menghubungkan data peserta didik dengan soal pada skala yang sama. Skala ini diperoleh dari nilai logaritma peluang peserta didik tersebut mengerjakan dengan benar. Dengan menggunakan pemodelan Rasch ini dapat dilihat hubungan

tingkat kemampuan peserta didik (*person ability*) dengan tingkat kesulitan soal (*item difficulty*). Dengan demikian dapat disimpulkan peserta didik dengan kemampuan tinggi akan mampu mengerjakan soal yang lebih mudah (Bond & Fox, 2007).

Analisis model Rasch mampu memberikan info secara keseluruhan, kualitas instrumen yang digunakan, kualitas respon peserta didik secara keseluruhan maupun interaksi antara responden dengan item soal (Chan, Ismail, & Sumintono, 2014). *Person measure* menunjukkan rata-rata skor nilai seluruh peserta didik dalam mengerjakan soal. Sumintono mengemukakan untuk mengukur reliabilitas, ditunjukkan dengan nilai *person reliability* dan *item reliability* (Sumintono & Widhiarso, 2015). Kriteria reliabilitas sebagai berikut: < 0,67 (lemah); 0,67-0,8 (cukup); 0,8-0,9 (bagus); 0,91-0,94 (bagus sekali); > 0,94 (istimewa).

Analisis model Rasch akan memberikan tingkat kesesuaian butir soal (*item fit*). *Item fit* akan menjelaskan apakah butir soal berfungsi dengan normal untuk melakukan pengukuran. Apabila diperoleh soal yang tidak fit maka hal ini menunjukkan adanya miskonsepsi pada soal sehingga berguna bagi guru untuk memperbaiki kualitas pengajarnya (Sumintono & Widhiarso, 2015).

Dalam analisis Rasch model kualitas kesesuaian butir dengan model disingkat dengan kesesuaian butir (*item fit*). Penentuan item fit dalam menurut Boone et al dalam Sumintono (2014) kriteria yang digunakan yaitu: (1) Nilai Outfit mean square (MNSQ) yang diterima = 0,5 < MNSQ < 1,5; (2) Nilai Outfit Z-standard (ZSTD) yang diterima = -2,0 < ZSTD < +2,0; (3) Nilai *Measure Correlation (Pr Mean Corr)*: 0,4 < *Pt Measure Corr* < 0,85.

Analisis data yang diperoleh dengan menggunakan *software Winsteps* sesuai dengan model Rasch apabila nilai jumlah kuadrat tengah (*mean square*) adalah 1,0 sedangkan nilai terstandarkannya (*Z-standardized value*) adalah 0,0. Pada analisis ini juga dapat diketahui soal yang terlalu mudah atau terlalu sulit dan responden yang *outlier*.

HASIL DAN PEMBAHASAN

Analisis data yang dihasilkan akan dibagi dalam beberapa ringkasan statistik yang memberikan petunjuk mengenai reabilitas dan validitas soal maupun responden. Hasil dari analisis Iteman dapat dilihat pada Tabel 1.

Tabel 1. Hasil Analisis Iteman

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics				
		Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
7	0-7 CHECK THE KEY C was specified, B works better	0.789	0.199	0.141	A	0.078	-0.584	-0.317	?
					B	0.078	0.386	0.210	
					C	0.789	0.199	0.141	
					D	0.056	-0.255	-0.125	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
8	0-8	0.333	0.549	0.423	A	0.300	-0.107	-0.082	*
					B	0.156	-0.602	-0.397	
					C	0.211	-0.063	-0.045	
					D	0.333	0.549	0.423	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	
9	0-9	0.578	0.467	0.370	A	0.244	-0.594	-0.434	*
					B	0.078	0.098	0.053	
					C	0.100	-0.061	-0.036	
					D	0.578	0.467	0.370	
					E	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

Dari Tabel 1, untuk menganalisisnya hal yang harus kita perhatikan adalah Prop. Correct yang menyatakan tingkat kesukaran butir, *Biser* merupakan indeks daya pembeda soal dengan menggunakan koefisien korelasi biserial sedangkan *Prop. Endorsing*= proporsi jawaban pada setiap option. Soal nomor 9 Prop. Correct = 0,578 hal ini menyatakan soal nomor 9 merupakan soal yang sedang, nilai bisernya 0,370 artinya soal ini mempunyai daya beda yang cukup, dilihat dari nilai *Prop. Endorsing* tiap option dipilih lebih dari 5% peserta didik, dengan demikian soal ini dapat digunakan.

Hal yang berbeda terlihat pada soal nomor 7, terdapat catatan kunci jawabannya *option C* namun B bekerja lebih baik. Jika dilihat dari point bisernya soal nomor 7 merupakan soal yang mempunyai daya beda rendah yaitu 0,141. Sementara itu nilai *Prop. Endorsing* pilihan C sebesar 0,789 artinya sebanyak 78,9% peserta didik memilih jawaban C dan jawaban tersebut merupakan jawaban yang benar. Nilai bisernya 0,141 dengan demikian soal ini kurang mampu membedakan peserta didik. Maka soal nomor 4 perlu diperbaiki option jawabannya.

Dari analisis Iteman ada beberapa soal yang perlu direvisi yaitu soal nomor 1, 2 dan 7. Item soal nomor 1 mempunyai *biser* = 0,835 dan *point biser* = 0,489 yang artinya soal nomor 1 tersebut jelek karena tidak dapat membedakan kemampuan responden. Soal nomor 1 dan soal nomor 7 terlalu sukar bagi responden sehingga soal ini perlu direvisi atau diganti.

Selain analisis tiap butir soal dalam analisis Iteman juga ditampilkan nilai statistik dari instrumen yang dikembangkan yang disajikan dalam Tabel 2.

Tabel 2. Statistik Iteman

Scale Statistics	
Scale:	0
N of Items	20
N of Examinees	90
Mean	10.789
Variance	18.078
Std. Dev.	4.252
Skew	-0.454
Kurtosis	-1.065
Minimum	3.000
Maximum	18.000
Median	12.000
Alpha	0.800
SEM	1.903
Mean P	0.539
Mean Item-Tot.	0.436
Mean Biserial	0.552

Dari Tabel 2 dapat diketahui *N of item* menunjukkan jumlah item soal yaitu sebanyak 20 soal, *N of examinees* menyatakan jumlah responden yaitu 90 peserta didik. Dari Tabel 2 dapat disimpulkan bahwa rata-rata peserta didik dapat menjawab 10,789 atau 10 soal. Dapat pula dikatakan skor nilai rata-ratanya 50 dari maksimal skor 100. *Variance* 18,078 menunjukkan varian dari distribusi skor responden yang memberikan gambaran tentang sebaran skor responden. *Skew* menunjukkan angka -0,454 yang berarti sebagian besar skor berada pada bagian atas. *Kurtosis* negatif menunjukkan distribusi yang landai (merata). *Variance*, standart deviasi, *skew* dan *kurtosis* menunjukkan sebaran skor responden, karena nilai *kurtosis* -1,065 yang berada diantara -2 dan 2 maka dapat disimpulkan bahwa sebaran skor nilai responden normal. Skor jawaban benar minimum 3 sedangkan maksimum jawaban benar 18 soal sedangkan mediannya 12 yang berarti skor nilainya 15-90

dengan rata-rata 50. *Alpha Cronbach* 0,80 berarti reliabilitas dari soal ini tinggi.

Pada Tabel 3 disajikan ringkasan dari analisis model Rasch, data diperoleh dari 75 responden dan 20 item soal. Responden menentukan satu jawaban dari empat pilihan jawaban yang disediakan. Selanjutnya data ini diolah dengan menggunakan *software winstep*.

Dari Tabel 3 diketahui bahwa jumlah data 1800 menghasilkan nilai *Chi-kuadrat* 1803,76 dengan derajat kebebasan 1691 ($p = 0,0281$) yang menunjukkan secara keseluruhan pengukuran memiliki signifikansi yang bagus hasilnya. Pada tabel yang mengukur pola jawaban responden diperoleh bahwa nilai *infit MNSQ* 0,95 dan *output MNSQ* 1,33 yang ekspektasinya 1, hal ini menunjukkan bahwa pola jawaban

responden pada instrumen bagus. Sedangkan nilai *infit ZSTD* - 0,2 dan *output ZSTD* - 0,2 yang ekspektasinya 0, hal ini juga menunjukkan bahwa secara keseluruhan pola jawaban responden sesuai dengan model. Apabila dilihat reliabilitas secara keseluruhan dari responden menunjukkan angka 0,70 yang artinya reliabilitas responden cukup. Reliabilitas responden ini terjadi karena responden yang digunakan hanya sebanyak 90 peserta didik. Sedangkan reliabilitas item soal 0,95 yang artinya istimewa. Interaksi responden dengan item soal sedang karena nilai *alpha Cronbach* 0,74 yang berarti cukup.

Pengujian instrumen pada item soalnya, terlihat nilai *infit MNSQ* 0,98 dan *output MNSQ* 1,33, Sedangkan nilai *infit ZSTD* -0,5 dan *output*.

Tabel 3. Ringkasan Statistik Instrumen: Responden dan Item soal

SUMMARY OF 90 MEASURED Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	10.2	20.0	.00	.55	.95	-.2	1.33	.2	
S.D.	3.7	.0	1.06	.04	.29	1.1	1.17	1.4	
MAX.	16.0	20.0	1.79	.66	1.65	2.3	6.02	3.4	
MIN.	3.0	20.0	-2.17	.51	.44	-2.2	.27	-2.0	
REAL RMSE	.58	TRUE SD	.88	SEPARATION	1.52	Person	RELIABILITY	.70	
MODEL RMSE	.55	TRUE SD	.90	SEPARATION	1.62	Person	RELIABILITY	.72	
S.E. OF Person MEAN = .11									
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00									
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .74									
SUMMARY OF 20 MEASURED Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	46.1	90.0	.00	.26	.98	-.5	1.33	.0	
S.D.	19.4	.0	1.24	.03	.21	1.6	1.14	2.3	
MAX.	72.0	90.0	2.56	.36	1.41	2.5	5.38	4.7	
MIN.	9.0	90.0	-1.70	.24	.62	-3.6	.57	-3.2	
REAL RMSE	.28	TRUE SD	1.21	SEPARATION	4.34	Item	RELIABILITY	.95	
MODEL RMSE	.27	TRUE SD	1.21	SEPARATION	4.55	Item	RELIABILITY	.95	
S.E. OF Item MEAN = .29									
UMEAN=.0000 USCALE=1.0000									
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00									
1800 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 1803.76 with 1691 d.f. p=.0281									
Global Root-Mean-Square Residual (excluding extreme scores): .3977									
Capped Binomial Deviance = .2176 for 1800.0 dichotomous observations									

Tabel 4. Perbandingan Hasil Analisis Iteman dan Rasch Model

No.	Komponen	Iteman	Rasch
1.	Reliabilitas	0,80	Person: 0,72 Item:0,95 Alpha Crobach: 0,74
2.	Validitas	0,93	0,93
3.	Daya Beda Soal	0,436	Kurva berdistribusi normal
4.	Tingkat kesukaran	0,539	6 separation
5.	Urutan tingkat kesukaran	7,20,18,19,3,13,17,14,6,9,10,16,11,4,5,15,12,8,2,1	7, 20,18,19,17,3,13,6,14,9,10,11,16,15,12,8,2,4,5,1.
6.	Item soal yang perlu direvisi	1,2,7	1, 4, 2,7

ZSTD 0 (ekspektasinya 0). Kedua hal ini menunjukkan bahwa reliabilitas item soal bagus, hal ini juga diperkuat dengan nilai reabilitasnya yaitu 0,95. Pengelompokan responden pada item soal cukup bagus yaitu 4,34 hal ini menunjukkan bahwa item soal ini mampu mengukur responden dengan kemampuan rendah sampai tinggi. Dengan demikian data aktual yang diperoleh pada penelitian ini cukup bagus sehingga dapat dilakukan analisis selanjutnya.

Sebaran responden dan item soal, dapat dilihat pada tabel *person-map-item*. Dari tabel tersebut terlihat hanya ada item soal yang outlier terlalu sulit sedangkan outlier mudah tidak ada. Terdapat dua soal yang outlier yaitu pada item soal interpretasi nomor 1 dan soal analisis nomor 5 karena pada kedua soal ini tidak ada responden yang dapat mengerjakan. Soal nomor 1 menanyakan bagaimana cara menggunakan kaliper yang benar. Sedangkan soal nomor 5 menanyakan faktor yang mempengaruhi gaya gesekan.

Responden yang *outlier* berkemampuan tinggi tidak ada terdapat responden yang outlier berkemampuan rendah. Responden tersebut identitasnya adalah 61, 68, 72, 75 dan 79. Kelima responden ini berasal dari SMK yang sama yaitu peserta didik yang berasal dari SMK kategori rendah. Hasil dari peserta didik ini selanjutnya dicermati dengan menggunakan data *scologram* ternyata responden ini mampu menjawab soal yang sulit namun tidak dapat menjawab soal yang mudah maka dapat diprediksi jika jawaban tersebut hasil dari *lucky guess*. Dengan demikian dapat diprediksi jika responden ini memang tidak bisa mengerjakan dengan baik.

Nilai logit rata-rata dari responden adalah 1,06 di atas 0,00 yang menunjukkan bahwa responden mempunyai kecenderungan dapat menjawab soal dengan benar. Nilai dari *separation* 4,34. Apabila digunakan rumus formula *person strata* yaitu H , sehingga nilai $H = [(4 * separation) + 1] / 3$ maka $H = [(4 * 4,34) + 1] / 3$, $H = 6,12$ atau dibulatkan menjadi 6. Hal ini menunjukkan bahwa responden terdiri dari enam kelompok saja. Semakin banyak *separation* artinya reabilitas respon cukup. Reabilitas responden ini terjadi karena responden yang digunakan hanya sebanyak 90 peserta didik. Sedangkan reliabilitas item soal 0,95 yang artinya istimewa. Interaksi responden dengan item soal sedang karena nilai *alpha Cronbach* 0,74 yang berarti cukup.

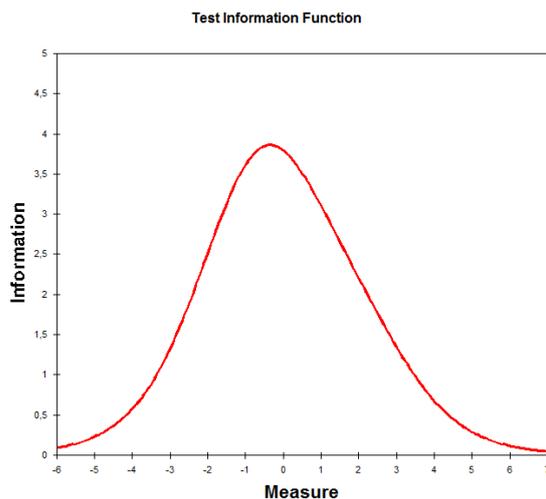
Pengujian instrumen pada item soalnya, terlihat nilai infit MNSQ 0,98 dan output MNSQ 1,33, Sedangkan nilai infit ZSTD -0,5 dan output ZSTD 0 (ekspektasinya 0). Kedua hal ini menunjukkan bahwa reliabilitas item soal bagus, hal ini juga diperkuat dengan nilai reabilitasnya yaitu 0,95. Pengelompokan responden pada item soal cukup bagus yaitu 4,34 hal ini menunjukkan bahwa item soal ini mampu mengukur responden dengan kemampuan rendah sampai tinggi. Dengan demikian data aktual yang diperoleh pada penelitian ini cukup bagus sehingga dapat dilakukan analisis selanjutnya.

Sebaran responden dan item soal, dapat dilihat pada tabel *person-map-item*. Dari tabel tersebut terlihat hanya ada item soal yang outlier terlalu sulit sedangkan outlier mudah tidak ada. Terdapat dua soal yang outlier yaitu pada item soal interpretasi nomor 1 dan soal analisis nomor 5 karena pada kedua soal ini tidak ada responden yang dapat mengerjakan. Soal nomor 1 menanyakan bagaimana cara menggunakan kaliper yang benar. Sedangkan soal nomor 5 menanyakan faktor yang mempengaruhi gaya gesekan.

Responden yang outlier berkemampuan tinggi tidak ada terdapat responden yang outlier berkemampuan rendah. Responden tersebut identitasnya adalah 61, 68, 72, 75 dan 79. Kelima responden ini berasal dari SMK yang sama yaitu peserta didik yang berasal dari SMK kategori rendah. Hasil dari peserta didik ini selanjutnya dicermati dengan menggunakan data *scologram* ternyata responden ini mampu menjawab soal yang sulit namun tidak dapat menjawab soal yang mudah maka dapat diprediksi jika jawaban tersebut hasil dari *lucky guess*. Dengan demikian dapat diprediksi jika responden ini memang tidak bisa mengerjakan dengan baik.

Nilai logit rata-rata dari responden adalah 1,06 di atas 0,00 yang menunjukkan bahwa responden mempunyai kecenderungan dapat menjawab soal dengan benar. Nilai dari *separation* 4,34. Apabila digunakan rumus *formula person strata* yaitu H , sehingga nilai $H = [(4 * separation) + 1] / 3$ maka $H = [(4 * 4,34) + 1] / 3$, $H = 6,12$ atau dibulatkan menjadi 6. Hal ini menunjukkan bahwa responden terdiri dari enam kelompok saja. Semakin banyak *separation* menunjukkan bahwa instrumen yang dibuat bagus. Semakin tinggi nilai *separasi butir*, semakin baik pengukuran yang dilakukan (Sumintono & Widhiarso, 2015, p 12).

Validitas responden juga dapat dilihat dari sebaran pola jawaban responden (di sebelah kiri) dengan tingkat kesulitan soal (di sebelah kanan). Nilai logit menunjukkan jika responden berada dibagian atas maka dia mampu menyelesaikan soal yang sulit (semakin keatas menunjukkan item soal semakin sulit). Identitas responden angka dan huruf yang menunjukkan asal sekolah, dimana M menunjukkan SMK St. Mikael (nomor 31-60), P untuk SMK Pancasila (nomor 61-90) dan R untuk SMK PGRI (nomor 1-30). Pola respon dari responden terhadap item soal cukup menarik. Ketiga SMK ini memberikan sebaran pada enam daerah yaitu mudah, sedang dan sulit. Dengan demikian item soal ini dapat digunakan untuk mengukur kemampuan berpikir kritis pada tiga sekolah tersebut, hal ini juga diperkuat dengan grafik tes *information function* yang diperoleh dari analisis *Winstep* pada Gambar 1.



Gambar 1. Sebaran Responden dan Item Soal

Dari Gambar 1 dapat dilihat sebaran tingkat kesukaran soal test kemampuan berpikir kritis. Nampak bahwa tidak semua aspek kemampuan berpikir kritis mempunyai tingkat kesukaran yang sama. Soal nomor 1 mengukur kemampuan interpretasi merupakan soal yang sulit bahkan soal ini masuk dalam outlier. Jika dilihat dari keenam aspek kemampuan berpikir kritis instrumen soal ini masih perlu revisi karena keenam aspek kemampuan berpikir kritis mempunyai tingkat kesukaran yang bervariasi.

Rata-rata logit item adalah 1,06 yang artinya soal dapat mengukur apa yang seharusnya diukur. Nilai ini menunjukkan bahwa kemungkinan responden dapat menjawab soal dengan benar memiliki peluang yang sama yaitu 50:50. Pada Gambar 1 terlihat bahwa item soal mem-

punyai range dari logit -3 sampai logit 3. Hal ini diperkuat dengan nilai separation 4,34 dan reliabilitas 0,95 yang memperkuat bahwa instrumen soal kemampuan berpikir kritis ini bagus. Dari Gambar 1 tentang grafik *test information function* nampak bahwa kurva berdistribusi normal, sesuai dengan model. Hal ini menunjukkan bahwa test kemampuan berpikir kritis yang dikembangkan dapat mengukur apa yang seharusnya diukur.

Aitem maupun instrument pengukuran dapat bersifat bias karena adanya perbedaan dalam responden. Pada penelitian ini responden berasal dari 3 SMK yang ada di Surakarta. Analisis model Rasch menampilkan hal ini dalam keberfungsian aitem differensial (*differential item functioning*, DIF). Hasil analisis DIF yang menunjukkan beberapa aitem terjangkit bias, dimana bisa diketahui dengan nilai probabilitas di bawah 5% (0,05) (Sumintono & Widhiarso, 2014). Dari grafik tersebut terlihat bahwa kemampuan berpikir kritis dipengaruhi oleh berbagai hal yang dapat diselidiki pada penelitian berikutnya. Tampilan grafik ini tidak dapat menyimpulkan dengan tegas bahwa kemampuan berpikir kritis peserta didik dari SMK M lebih baik dibandingkan peserta didik SMK yang lain. Dengan demikian justru dapat disampaikan bahwa kemampuan berpikir kritis peserta didik tidak tergantung dari SMK asalnya.

Tabel keberfungsian *DIF* membandingkan bagaimana responden dari 3 sekolah memberikan respon pada tiap item soal. Pada soal interpretasi I1 dan analisa A5 terlihat bahwa responden dari sekolah M mudah mengerjakannya sedangkan bagi responden sekolah R item soal ini sulit. Sementara itu pada item soal analisa A4 responden dari sekolah P dapat mengerjakan dengan baik sedangkan responden dari sekolah M posisinya berimpit yang menunjukkan pemahaman mereka terhadap item soal ini sama sedangkan dari responden R tidak dapat menganalisis dengan baik. Pada soal evaluasi E1, IN1, P2, P3,R2,R3 dan I3 responden dari sekolah R dapat mengerjakan dengan baik dibanding reponden dari sekolah M dan R. Pada item soal analisa 6 dan regulasi diri 1 nampak responden M mengalami penurunan kemampuan mengerjakan soal yang hal ini berkebalikan dengan responden kelompok M sementara itu kelompok responden P terlihat stabil. Dari sebaran keenam aspek kemampuan berpikir kritis tersebut dapat dinyatakan bahwa rata-rata kemampuan kelompok responden M lebih baik

dibandingkan dengan R dan P. Sedangkan kemampuan kelompok responden P lebih baik dibandingkan kelompok R. Kelompok P menunjukkan kestabilan dalam menjawab soal, hal ini berbeda dengan kelompok R yang kemampuannya tak dapat diprediksikan. Ada kemungkinan kelompok R ini menjawab dengan menebak.

Urutan tingkat kesukaran dari yang mudah ke sulit yaitu dari *scalogram* dapat dilihat pula asal sekolah responden yang menjawab dengan dengan nilai tertinggi 5 responden teratas berasal dari sekolah dengan kode M. Apabila dikelompokkan dalam 3 kategori 30 responden teratas terdiri dari 23 peserta didik sekolah M dan 7 peserta didik sekolah P. Kelompok responden tengah terdiri 3 peserta didik dari sekolah R, 20 peserta didik P dan 7 peserta didik sekolah M. Kelompok responden 30 kebawah terdiri dari 27 peserta didik sekolah R dan 3 peserta didik P. Ketiga sekolah ini apabila dilihat dari nilai rangking ujian nasional tahun 2016/2017 urutannya adalah M, P dan R apabila dibandingkan dengan tes kemampuan berpikir kritis maka memberikan data yang sama. Ada beberapa faktor yang mungkin mempengaruhinya yaitu pemilihan responden, input dari peserta didik yang berbeda dan kemampuan peserta didik tersebut.

Analisis iteman dan analisis pemodelan Rasch apabila digabungkan akan memberikan data yang saling melengkapi. Analisis iteman menganalisis soal secara keseluruhan sedangkan analisis pemodelan Rasch mampu menganalisis hubungan antara item soal dengan responden. Pada analisis iteman responden dengan skor yang sama dianggap mempunyai kemampuan yang sama, sedangkan pada pemodelan Rasch responden dengan skor yang sama dapat dilihat tingkat kemampuannya.

Dalam analisis Rasch model dapat mengetahui *person fit*. Penentuan person fit juga menggunakan kriteria yang sama dengan item fit. Sebagian besar responden yang berasal dari sekolah R merupakan responden yang *unfit*. Untuk mengetahui lebih jauh tentang responden tersebut perlu didukung data dari analisis *scalograms*. Responden dengan menunjukkan ketidakstabilan dalam mengerjakan item soal, peserta didik ini mampu mengerjakan soal mudah namun ada soal mudah yang lain yang tidak bisa dikerjakan sementara itu soal yang sulit dia mampu mengerjakannya dengan demikian dapat diprediksi jawaban betul peserta didik tersebut karena menebak saja.

Empat urutan pertama sama urutan berikutnya berbeda. Namun soal yang dianggap paling sulit tetap sama yaitu soal nomor satu. Namun setelah urutan kelima terjadi beberapa pergeseran, hal ini dapat terjadi karena dalam analisis Rasch analisis item soal juga memperhatikan pola responden. Walaupun demikian pada analisis Iteman dan analisis pemodelan Rasch diperoleh data yang hampir sama.

Pada kedua analisis ini ada item soal yang perlu direvisi. Dari analisis Iteman soal yang perlu direvisi yaitu nomor 1,2 dan 7 sedangkan menurut Rasch model nomor soal yang perlu direvisi yaitu soal nomor 1, 4, 2 dan 7. Ada satu nomor yang berbeda yaitu nomor 4. Menurut analisis Iteman soal nomor 4 merupakan soal yang mempunyai daya beda soal dan taraf kesukaran yang baik. Sedangkan dalam analisis Rasch memperhatikan *item infit*, hal yang perlu dicermati yaitu *MNSQ outfit*, *ZSTD outfit* dan *PT-Measure Corr*. Item tersebut tidak memenuhi kriteria karena mempunyai *MNSQ outfit* lebih dari 1,5; *ZSTD outfit* lebih dari 2 atau kurang dari - 2 dan *PT-Measure Corr* kurang dari 0,4.

Dari analisis iteman dan analisis pemodelan Rasch dapat diketahui bahwa instrumen kemampuan berpikir kritis yang sudah dikembangkan memiliki reliabilitas dan validitas yang baik. Sehingga instrumen ini dapat digunakan untuk penelitian lebih lanjut.

SIMPULAN

Artikel ini menyajikan riset tentang pengembangan dan validasi instrumen kemampuan berpikir kritis. Enam aspek kemampuan berpikir kritis yang dikemukakan oleh Facione dijadikan sebagai dasar penyusunan indikator instrumen. Keenam aspek tersebut meliputi interpretasi, analisis, evaluasi, inference, penjelasan dan regulasi diri. Responden terdiri dari 90 peserta didik dari 3 SMK di Surakarta (SMK M, SMK R dan SMK P) dipilih secara acak mengerjakan instrumen penelitian.

Hasil analisis dengan pemodelan iteman diperoleh nilai *Alpha Cronbach* 0,8 berarti reliabilitas dari soal ini tinggi. Sedangkan dengan model Rasch *Alpha Cronbach* 0,74 berarti reliabilitas antara person dan item cukup, sedangkan reliabilitas item 0,92 yang berarti item instrumen sangat bagus. Validitas item 0,93 menunjukkan soal dapat mengukur dan diperkuat dengan separation 4,34 item soal ini mempunyai sebaran respon yang baik. Instrumen

yang dikembangkan dapat digunakan untuk penelitian selanjutnya.

DAFTAR PUSTAKA

- Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan* (2nd ed.). Jakarta: Bumi Aksara.
- Azwar, S. (2015). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Bond, T. G., & Fox, C. M. (2007). Applying the rasch model: fundamental measurement in the human sciences, Second Edition. *Journal of Educational Measurement*, 2(2nd), 360. <https://doi.org/10.1111/j.1745-3984.2003.tb01103.x>
- Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics. *Procedia - Social and Behavioral Sciences*, 129, 133–139. <https://doi.org/10.1016/j.sbspro.2014.03.658>
- Facione, P. A. (2011a). *Critical thinking : What it is and why it counts*.
- Facione, P. A. (2011b). Critical thinking: What it is and why it counts. In *Insight Assessment*.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). An Introduction to educational design research. *East*, 129.
- Hadam, S., Rahayu, N., & Nur Ariyadi, A. (2018). *10 Langkah Revitalisasi SMK*. Jakarta: Direktorat Pembinaan Sekolah Menengah Kejuruan Direktorat Jenderal Pendidikan Dasar Dan Menengah Kementerian Pendidikan dan Kebudayaan.
- Kompas. (2017, November 14). Lulusan vokasi menganggur. *Kompas*. Jakarta.
- Lai, E. R. (2011). Critical thinking : A literature review research Report, (June).
- Morganna, K., Pinto, P., Federal, U., & Grande, R. (2007). Development and validation of an instrument for measuring patient satisfaction with physical therapy, 369–376.
- Ngalimun. (2018). *Evaluasi dan penilaian pembelajaran*. Yogyakarta: Parama Ilmu.
- Peter, A., & Facione, P. A. (1989). The California critical thinking skills test -- college level technical report # 4 Interpreting the CCTST , Group Norms , and Sub-Scores.
- Schwab, K. (2016). The future of jobs employment, skill and workforce strategy for the fourth industry revolution.
- Starkey, L. (2009). *Critical thinking skills success tes kemampuan berpikir kritis dalam 20 menit*. Yogyakarta: Bookmarks.
- Sudjana. (2005). *Penilaian hasil proses belajar mengajar*. Bandung: PT Remaja Rosdakarya.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial* (2nd ed.). Cimahi: Trim Komunikata Publishing House.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Cimahi: Trim Komunikata Publishing House.
- Trilling, B., & Fadel, C. (2009). *21 st century skills*. San Francisco: Jossey- Bass.
- Zhou, Q., Huang, Q., & Tian, H. (2013). Developing students ' critical thinking skills by task-based learning in chemistry experiment teaching, 4(12), 40–45.