

PENGEMBANGAN TES BERPIKIR KRITIS DENGAN PENDEKATAN *ITEM RESPONSE THEORY*

¹⁾Fajrianti, ²⁾Wiwin Hendriani, ³⁾Berlian Gressy Septarini

^{1, 2, 3)}Fakultas Psikologi Universitas Airlangga

¹⁾fajrianti@psikologi.unair.ac.id, ²⁾wiwin.hendriani@psikologi.unair.ac.id,

³⁾gressy.septarini@psikologi.unair.ac.id

Abstrak

Penelitian ini bertujuan untuk menghasilkan sebuah alat ukur (tes) berpikir kritis yang valid dan reliabel untuk digunakan, baik dalam lingkup pendidikan maupun kerja di Indonesia. Tahapan penelitian dilakukan berdasarkan tahap pengembangan tes menurut Hambleton dan Jones (1993). Kisi-kisi dan pembuatan butir didasarkan pada konsep dalam tes *Watson-Glaser Critical Thinking Appraisal* (WGCTA). Pada WGCTA, berpikir kritis terdiri dari lima dimensi yaitu *Inference*, *Recognition Assumption*, *Deduction*, *Interpretation* dan *Evaluation of arguments*. Uji coba tes dilakukan pada 1.453 peserta tes seleksi karyawan di Surabaya, Gresik, Tuban, Bojonegoro, Rembang. Data dikotomi dianalisis dengan menggunakan model IRT dengan dua parameter yaitu daya beda dan tingkat kesulitan butir. Analisis dilakukan dengan menggunakan program statistik Mplus versi 6.11 Sebelum melakukan analisis dengan IRT, dilakukan pengujian asumsi yaitu uji unidimensionalitas, independensi lokal dan *Item Characteristic Curve* (ICC). Hasil analisis terhadap 68 butir menghasilkan 15 butir dengan daya beda yang cukup baik dan tingkat kesulitan butir yang berkisar antara -4 sampai dengan 2.448. Sedikitnya jumlah butir yang berkualitas baik disebabkan oleh kelemahan dalam menentukan *subject matter experts* di bidang berpikir kritis dan pemilihan metode skoring.

Kata kunci: Pengembangan tes, berpikir kritis, item response theory

DEVELOPING CRITICAL THINKING TEST UTILISING ITEM RESPONSE THEORY

¹⁾Fajrianti, ²⁾Wiwin Hendriani, ³⁾Berlian Gressy Septarini

^{1, 2, 3)}Fakultas Psikologi Universitas Airlangga

¹⁾fajrianti@psikologi.unair.ac.id, ²⁾wiwin.hendriani@psikologi.unair.ac.id,

³⁾gressy.septarini@psikologi.unair.ac.id

Abstract

The present study was aimed to develop a valid and reliable instrument in assessing critical thinking which can be implemented both in educational and work settings in Indonesia. Following the Hambleton and Jones's (1993) procedures on test development, the study developed the instrument by employing the concept of critical thinking from *Watson-Glaser Critical Thinking Appraisal* (WGCTA). The study included five dimensions of critical thinking as adopted from the WGCTA: *Inference*, *Recognition Assumption*, *Deduction*, *Interpretation* dan *Evaluation of arguments*. 1453 respondents from Surabaya, Gresik, Tuban, Bojonegoro and Rembang were used for trailing the test. The dichotomous data were analyzed using the Item Response Theory with two parameter logistic model using statistical program Mplus ver. 6.11. Several assumptions were tested prior the IRT analysis; the test of unidimensionality, local independency and *Item Characteristic Curve* (ICC). Amongst 68 items only 15 items had good discrimination parameter. Difficulty item level ranged from - 4.95 to 2.448. The study was limited in producing high number of qualified items due to its failure in finding *subject matter experts* in critical thinking area and inadequate choice in scoring method.

Keywords: test development, critical thinking, Item response theory

Pendahuluan

Pemecahan masalah merupakan kegiatan rutin yang berlangsung sepanjang kehidupan manusia. Efektifitas pemecahan masalah tergantung pada banyak faktor. Walaupun demikian, para ahli meyakini bahwa berpikir kritis merupakan salah satu faktor yang penting dalam pemecahan masalah baik di bidang pendidikan maupun bidang kehidupan lainnya. Bahkan, dalam beberapa tahun terakhir, penggunaan konstruk berpikir kritis sebagai prediktor keberhasilan di dunia pendidikan maupun di dunia kerja semakin banyak dilakukan (Wagner, 2002).

Berpikir kritis merupakan keterampilan penting yang dibutuhkan dalam dunia kerja di abad 21 (*Essential Skills for the 21st Century Workplace*, 2014). Keterampilan ini bahkan menduduki urutan pertama dalam daftar keterampilan yang dibutuhkan. Keterampilan komunikasi, kolaborasi, kesadaran global (*global awareness*), penguasaan teknologi, keterampilan dalam hidup dan karir, kemampuan belajar dan inovasi membutuhkan fondasi berpikir kritis yang baik.

Berpikir kritis dipandang sebagai hasil (*outcome*) dari pembelajaran di perguruan tinggi. Hal ini tidak hanya menjadi isu di pihak eksternal atau lingkungan masyarakat namun juga didalam perguruan tinggi itu sendiri. Isu utama yang muncul adalah tentang bagaimana akuntabilitas dari hasil pembelajaran berpikir kritis tersebut (Stassen, Herington, Henderson, 2011). Hal ini terkait dengan hasil survey yang dilakukan oleh Association of American Colleges and Universities (AAC&U) pada tahun 2008 yang menyatakan bahwa 73% dari karyawan menginginkan agar perguruan tinggi memberikan lebih banyak penekanan pada berpikir kritis dan penalaran analitis. Bahkan, survei AAC&U di tahun 2009 menunjukkan bahwa 74% responden menyatakan bahwa berpikir kritis merupakan inti dari tujuan pembelajaran dalam program pendidikan di kampus (Stassen dkk., 2011).

Pentingnya konstruk berpikir kritis dalam dunia pendidikan maupun pekerjaan memang tidak lagi diperdebatkan. Meskipun demikian, pendefinisian maupun bagaimana

konstruk berpikir kritis tersebut diukur masih menjadi perdebatan para ahli di bidang Psikologi, Filsafat maupun Pendidikan. Hal ini terkait dengan beragamnya definisi dan pengukuran konstruk berpikir kritis yang ada saat ini (Wagner, 2002; Stassen dkk., 2011).

Terdapat berbagai definisi tentang berpikir kritis, di antaranya Sternberg (1986) yang menyatakan bahwa berpikir kritis adalah proses mental, strategi dan representasi yang digunakan individu untuk memecahkan, membuat keputusan dan mempelajari konsep baru. Berpikir kritis merupakan investigasi yang bertujuan untuk mengeksplorasi situasi, fenomena, pertanyaan atau masalah untuk menjadi hipotesis atau kesimpulan melalui pengintegrasian seluruh informasi yang tersedia sehingga memiliki justifikasi yang meyakinkan (Kurfiss, 1988). Berpikir kritis mencakup kemampuan berpikir yang masuk akal (*reasonable*) dan reflektif yang berfokus pada keputusan tentang apa yang akan dipercaya atau dilakukan (Noris dan Ennis, 1989).

Angelo dan Cross (1995) menyebutkan berpikir kritis sebagai kegiatan berpikir tingkat tinggi, yang meliputi kegiatan menganalisis, mensintesis, mengenali permasalahan dan pemecahannya, menyimpulkan serta mengevaluasi. Kegiatan ini dilakukan terhadap berbagai informasi yang didapat dari hasil observasi, pengalaman, refleksi, di mana hasil proses ini digunakan sebagai dasar individu saat mengambil tindakan (Walker dan Finney, 1996). Sedikit berbeda dari penjelasan sebelumnya, Hossoubah (2007) mendefinisikan berpikir kritis sebagai kemampuan memberi alasan secara terorganisasi dan mengevaluasi kualitas suatu alasan secara sistematis.

Hasil penelitian pendahuluan menunjukkan bahwa perdebatan mengenai definisi berpikir kritis lebih terletak pada deskripsi tentang indikator daripada perbedaan yang lebih bersifat fundamental. Hal ini didukung oleh laporan hasil penelitian yang ditulis dalam *Critical Thinking: A Literature Review* (Lai, 2011). Dalam *Literature Review* ini dinyatakan bahwa di antara berbagai perbedaan

an mengenai definisi berpikir kritis terdapat kesepakatan bahwa berpikir kritis merupakan kemampuan berpikir yang terdiri dari komponen: keterampilan dalam menganalisis argumen, membuat kesimpulan baik secara induktif maupun deduktif, mengevaluasi dan membuat keputusan atau memecahkan masalah (Lai, 2011).

Pengukuran konstruk berpikir kritis di Indonesia cukup beragam, terutama terkait dengan konteks pengukurannya. Selama ini juga terdapat perdebatan tentang apakah berpikir kritis merupakan konstruk yang spesifik pada setiap *setting* atau bersifat general. Sejauh pengamatan peneliti selama ini, pengukuran berpikir kritis di Indonesia lebih sering menggunakan tes-tes yang dikembangkan di dunia barat. Misalnya, *California Critical Thinking Skills Test* dan *Cornell Critical Thinking Tests* yang digunakan tanpa adanya pengujian terlebih dahulu, apakah tes ini mengandung bias budaya atau tidak. Selain itu pengembangan pengukuran berpikir kritis di Indonesia lebih banyak dilakukan pada *setting* pendidikan matematika dan fisika. Sampai saat ini, di Indonesia belum ada tes berpikir kritis yang terstandar dikembangkan dengan menggunakan *setting* yang general/umum, terutama dengan menggunakan pendekatan *Item Response Theory* (IRT).

IRT adalah sebuah model probabilitas yang berusaha menjelaskan hubungan antara respon seseorang terhadap sebuah butir dengan variabel laten (kemampuan/*ability* atau sifat/*trait*) yang diukur oleh tes tersebut. Dalam hal ini, respon atau kinerja peserta tes merupakan hal yang dapat diamati (*observable*) sedangkan sifat atau kemampuan merupakan sesuatu yang tidak tampak (*unobservable*) yang mendasari kinerja pada tes tersebut (Embretson & Reise, 2000). Variabel laten (misalnya kemampuan) dalam IRT disebut sebagai theta (θ) dimana semakin tinggi θ yang dimiliki seseorang maka semakin tinggi pula probabilitasnya untuk menjawab soal dengan benar (Baker, 2001). Probabilitas seorang peserta tes Berpikir Kritis untuk menjawab benar sebuah butir merupakan fungsi dari tingkat kemampuan berpikir kritis yang dimiliki dengan tingkat ke-

sulitan dari butir tersebut. Berikut merupakan model matematis dari IRT dengan 2 parameter.

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$$

$i = 1, 2, \dots, n$

Keterangan :

- θ : tingkat kemampuan peserta tes
- P_i : probabilitas peserta tes yang memiliki kemampuan θ menjawab benar soal i
- a_i : indeks daya pembeda soal ke i
- b_i : indeks kesulitan soal ke- i
- e : bilangan natural yang nilainya mendekati 2,718
- n : banyaknya soal tes
- D : faktor penskalaan yang harganya 1,7.

Lebih lanjut, kelemahan utama dari teori tes klasik adalah hasil pengukurannya yang bersifat *group dependent* dan *item dependent*. Parameter tingkat kesulitan soal dalam teori tes klasik tergantung pada peserta yang menjawab soal tersebut. Artinya, sebuah soal akan menjadi soal yang mudah jika dijawab oleh peserta yang pandai, namun soal tersebut sekaligus menjadi soal yang sulit jika dijawab oleh peserta yang kurang pandai. Demikian pula pada parameter kemampuan peserta, seorang peserta akan menjadi pandai saat mengerjakan soal yang mudah namun sekaligus menjadi kurang pandai saat mengerjakan soal yang sulit. Dengan demikian, tidak pernah diketahui tingkat kemampuan yang sebenarnya dari peserta tes. Seperti halnya juga tidak diketahui apakah sebuah soal memiliki tingkat kesulitan yang tinggi atau rendah, karena tergantung pada kemampuan peserta yang menjawabnya. Pada IRT hal yang demikian tidak terjadi, karena kita benar-benar dapat mengukur tingkat kemampuan seseorang dengan lebih akurat, begitu pula dalam mengestimasi parameter soal/butir.

Berdasarkan uraian latar belakang yang telah dipaparkan, dapat disimpulkan dua hal yang menjadi dasar penelitian ini. Pertama, pentingnya konstruk berpikir praktis dalam memprediksi keberhasilan seseorang dalam

pendidikan ataupun pekerjaan memunculkan kebutuhan akan tersedianya tes berpikir praktis yang valid dan reliabel sehingga dapat memberikan hasil ukur yang akurat. Tes-tes berpikir kritis yang telah ada di Indonesia lebih banyak merupakan penerjemahan dari tes-tes yang beredar di Barat (Amerika dan Eropa) tanpa didahului pengujian tentang bias budayanya. Hal ini akan berdampak pada keakuratan hasil ukurnya. Kedua, pengembangan tes sendiri di Indonesia masih lebih banyak menggunakan pendekatan Teori Tes Klasik (*Classical Test Theory*) yang memiliki sejumlah kelemahan. Oleh karena itu, penelitian ini bermaksud menindaklanjuti kebutuhan tersebut dengan mengembangkan tes berpikir kritis yang valid dan reliabel dengan menggunakan pendekatan IRT. Tujuan penelitian ini adalah mengembangkan tes Berpikir Kritis yang valid dan reliabel untuk digunakan dalam lingkungan pendidikan dan kerja di Indonesia.

Metode Penelitian

Tipe penelitian ini adalah penelitian survei, yaitu pengumpulan data dari kelompok masyarakat, yang datanya kemudian dianalisis untuk mengembangkan tes berpikir kritis yang dapat digunakan oleh kelompok masyarakat tersebut.

Tahapan penelitian yang dilakukan berdasarkan tahap pengembangan tes menurut Hambleton dan Jones (1993), yaitu:

Menyiapkan Spesifikasi Tes (*Preparation of Test Specifications*)

Tahap pertama ini diawali dengan proses identifikasi perilaku yang merepresentasikan konstruk atau domain perilaku. Pada tahap ini, pengembang tes perlu menentukan satu atau lebih jenis perilaku yang diyakini merupakan manifestasi dari konstruk yang akan diukur, kemudian merumuskannya ke dalam bentuk butir yang dapat mengungkap perilaku tersebut.

Terdapat beberapa cara dalam menentukan domain perilaku dari sebuah konstruk, pada penelitian ini peneliti memilih

cara berdasarkan tinjauan terhadap hasil penelitian (*Review of research*) tentang tes berpikir kritis yang telah ada saat ini. Berdasarkan tinjauan terhadap berbagai penelitian di bidang tes berpikir kritis, peneliti memutuskan untuk menggunakan konsep yang digunakan dalam tes *Watson-Glaser Critical Thinking Appraisal* (WGCTA). Pada WGCTA, berpikir kritis terdiri dari lima dimensi yaitu *Inference*, *Recognition Assumption*, *Deduction*, *Interpretation* dan *Evaluation of arguments*. Berdasarkan kelima dimensi tersebut peneliti membuat *blue print* sebagaimana tertuang dalam Tabel 1.

Tabel 1. *Blue Print* Tes

Dimensi	Jumlah Butir
Kesimpulan (<i>Inference</i>): Kemampuan dalam menilai tingkat probabilitas ketepatan/kebenaran sebuah kesimpulan berdasarkan informasi yang tersedia.	30
Pengenalan asumsi (<i>Recognition Assumption</i>) Kemampuan dalam mengidentifikasi asumsi yang tersirat pada sebuah pernyataan	36
Deduksi (<i>Deduction</i>) Kemampuan dalam menentukan apakah kesimpulan dibuat secara logis berdasarkan informasi yang tersedia	20
Interpretasi (<i>Interpretation</i>) Kemampuan dalam menilai sebuah bukti (<i>evidence</i>) dan membuat keputusan apakah generalisasi / kesimpulan yang dihasilkan dijamin berdasarkan data yang tersedia	20
Evaluasi argumen (<i>Evaluation of arguments</i>) Kemampuan dalam mengevaluasi kekuatan dan relevansi sebuah argument terkait dengan sebuah isu atau masalah tertentu	30

Menyiapkan Butir-Butir Tes (*Preparation of The Test Item Pool*)

Peneliti menggunakan berbagai sumber literatur untuk mengembangkan skena-

rio yang akan digunakan sebagai batang soal. Selanjutnya berdasarkan masing-masing batang soal dibuat 4-6 buah butir. Jumlah butir yang dibuat sesuai dengan yang tertera di *blue print*. Setelah itu dilakukan review terhadap soal yang telah dibuat. Review dilakukan oleh para expert judges yang memahami tentang konsep tes berpikir kritis.

Uji Coba Butir Tes di Lapangan (*Field Testing The Items*)

Uji coba dilakukan untuk mengetahui apakah instruksi tes dapat dipahami dengan baik serta pernyataan dalam butir tidak memiliki pengertian yang ambigu. Kegiatan ini melibatkan 10 orang mahasiswa S1 Psikologi yang sedang mengikuti magang di Lembaga Pengkajian dan Psikologi Terapan (LP3T) Unair. Pemilihan subjek tersebut didasarkan pertimbangan bahwa mereka telah terbiasa mengenal dan memahami tes-tes Psikologi yang digunakan dalam kegiatan seleksi dan promosi karyawan.

Revisi Butir Tes (Revision of the Test Items)

Berdasarkan masukan dari 10 mahasiswa pada uji coba, dilakukan revisi yang terkait dengan administrasi dan kalimat pada butir. Pada tahap ini tes dibuat menjadi dua bagian yang paralel, sehingga masing-masing bagian tes hanya berisi separuh dari jumlah butir yang terdapat dalam setiap dimensi. Hal ini dilakukan agar peserta tes tidak merasa jenuh dan kehilangan konsentrasi karena disebabkan oleh banyaknya soal yang harus dikerjakan. Setelah dilakukan revisi, dilakukan pengujian di lingkungan internal Fakultas Psikologi dengan menggunakan subyek 30 mahasiswa S1 yang mengikuti mata kuliah Psikologi Industri dan Organisasi dan 38 mahasiswa peserta mata kuliah Metode Penelitian Kualitatif. Hasil dari uji coba ini dilakukan analisis dengan menggunakan *software* ITEMAN untuk mengetahui tingkat kesulitan, daya beda dan fungsi masing-masing distraktor. Berdasarkan hasil analisis tersebut dirakit tes berpikir kritis format final yang hanya berisi butir dengan kualitas yang baik.

Pengembangan Tes (*test development*)

Proses berikutnya adalah melakukan pengambilan data di lapangan. Data diperoleh dengan menggunakan subjek yang mengikuti tes seleksi karyawan Semen Gresik dan Angkasa Pura. Jumlah seluruh subjek adalah 1.453 orang.

Lebih lanjut, analisis data dilakukan dengan pendekatan Item Response Theory, menggunakan program Mplus versi 6.00. Hasil dari penelitian ini berupa dokumen berupa hasil analisis butir yang akan memberikan estimasi parameter butir yaitu tingkat kesulitan dan daya beda butir. Hasil analisis juga akan memberikan data tentang tingkat kemampuan berpikir kritis dari masing-masing subjek.

Hasil Penelitian dan Pembahasan

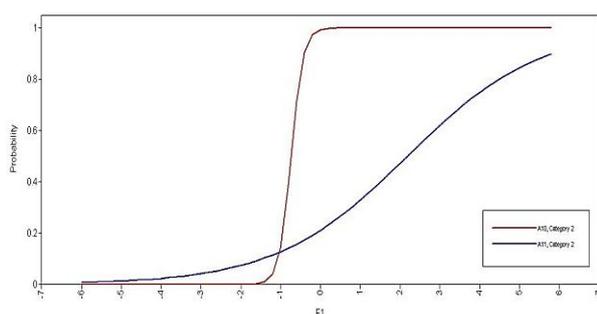
Sebelum melakukan analisis butir dengan menggunakan IRT, dilakukan uji asumsi yaitu unidimensionalitas, independensi lokal dan *Item Characteristic Curve* (ICC) (Hambleton dkk., 1991; Baker, 2001). Pengujian asumsi unidimensionalitas dilakukan dengan menggunakan CFA satu faktor dengan metode estimasi yaitu *robust weighted least squares* (WLS/WLSMV) yang sesuai dengan jenis data kategori/dikotomi. Pengujian dilakukan dengan *software* statistic Mplus versi 6.11.

Berdasarkan hasil pengujian CFA, diketahui bahwa subtes 1 sesuai dengan model satu faktor dengan nilai RMSEA = 0,026. Subtes 2 juga sesuai dengan model satu faktor dengan nilai RMSEA = 0,048. Subtes 3, datanya sesuai dengan model satu faktor dengan nilai RMSEA = 0,070. Subtes 4 juga sesuai dengan model satu faktor dengan nilai RMSEA = 0,054. Subtes 5 tidak sesuai dengan model satu faktor dengan nilai RMSEA = 0,111. Menurut MacCallum (dalam Hooper dkk., 2008) nilai RMSEA < 0,08 menunjukkan model yang *good fit*. Nilai RMSEA > 0.1 menunjukkan model yang *poor fit*. Artinya, dalam hal ini model subtes 5 tidak sesuai dengan data, sehingga subtes 5 bukan merupakan model satu faktor.

Hambleton dkk. (1991) mengemukakan bahwa jika asumsi unidimensionalitas dipenuhi maka asumsi independensi lokal juga terpenuhi. Berdasarkan hal tersebut maka keempat subtes pada tes berpikir kritis memenuhi asumsi independensi lokal.

Analisis butir tes kemampuan berpikir kritis dilakukan dengan menggunakan *software* statistik Mplus versi 6.11 Metode estimator yang digunakan dalam analisis ini adalah *Robust Maximum Likelihood* (MLR). Berikut adalah hasil dari analisis tersebut yang dilakukan pada setiap subtes.

Berdasarkan hasil analisis pada subtes 1 (kemampuan menilai ketepatan sebuah kesimpulan) diketahui bahwa hanya terdapat dua butir dengan daya beda yang lebih besar dari 0,3 yaitu butir nomor 10 dan 11. Kedua butir tersebut memiliki daya beda yang cukup baik, artinya kedua butir tersebut dapat membedakan antara individu yang memiliki kemampuan yang rendah dengan yang tinggi pada kondisi dimana kemampuan (*theta*) sama dengan tingkat kesulitan butir (b). Dengan demikian, kedua butir tersebut mampu membedakan antara individu yang memiliki kemampuan tinggi dalam menilai ketepatan sebuah kesimpulan dengan individu yang berkemampuan rendah. ICC dari butir nomor 10 dan 11 disajikan dalam grafik pada Gambar 1.



Gambar 1. ICC Subtes 1

Nilai daya beda yang rendah ditunjukkan dengan bentuk ICC yang datar, hal tersebut menunjukkan bahwa probabilitas menjawab benar antara individu dengan kemampuan yang tinggi dengan individu yang berkemampuan rendah adalah sama (Baker, 2001).

Nilai daya beda yang bertanda positif menunjukkan bahwa probabilitas menjawab benar butir tersebut meningkat seiring dengan meningkatnya taraf kemampuan (Baker, 2001). Sebaliknya, daya beda yang bernilai negatif menunjukkan bahwa probabilitas menjawab benar butir tersebut, menurun seiring dengan meningkatnya kemampuan dari tingkat yang rendah ke tingkat yang lebih tinggi (Baker, 2001). Butir yang memiliki daya beda negatif merupakan butir yang buruk. Hal tersebut dapat terjadi karena kesalahan pada saat penulisan butir, dimana terdapat kesalahan informasi yang membuat individu dengan kemampuan yang tinggi gagal memahami maksud dari butir tersebut sehingga salah dalam menjawab butir (Baker, 2001).

Tabel 2. Hasil Estimasi Parameter Butir Subtes 1

No. Butir	Daya Beda (a)	Tingkat Kesulitan (b)
1.	0,013	3,145
2.	-0,023	0,458
3.	-0,009	1,385
4.	0,103	1,345
5.	0,094	-0,200
6.	0,065	1,555
7.	-0,271	1,570
8.	0,042	1,067
9.	-0,054	2,430
10.	3,981	-4,959
11.	0,355	1,322
12.	0,176	-0,942
13.	-0,074	2,104
14.	-0,074	0,435
15.	0,063	1,065

Berdasarkan tingkat kesulitan butir dapat diketahui bahwa butir nomor 1 merupakan butir dengan tingkat kesulitan tertinggi sedangkan butir nomor 10 merupakan butir yang paling mudah. Tingkat kesulitan butir umumnya berada pada rentang -3 sampai dengan 3. Nilai kesulitan butir yang berada di luar rentang tersebut menunjukkan bahwa butir tersebut bermasalah yang mungkin terkait dengan isi maupun

penulisan butir (Sudol & Studer, 2010). Tingkat kesulitan item yang berada di luar rentang -3 sampai dengan 3 tidak sesuai dengan rentang kemampuan peserta yang umumnya berada pada rentang -3 sampai dengan 3 dengan nilai rata-rata 0 dan varians 1.

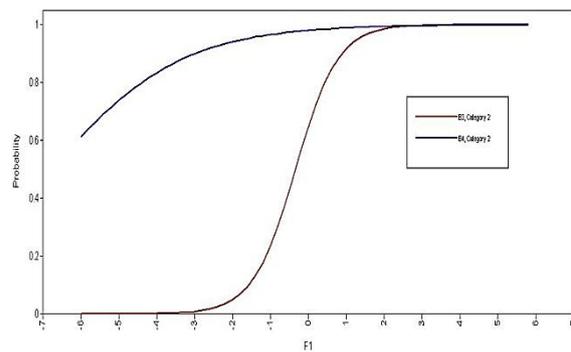
Tabel 3. Hasil Estimasi Parameter Butir Subtes 2

No. Butir	Daya Beda (a)	Tingkat Kesulitan (b)
1.	-1,281	0,082
2.	-0,577	2,448
3.	1,054	-0,609
4.	0,340	-3,920
5.	-0,150	-1,918
6.	-0,026	1,010
7.	0,065	0,032
8.	0,124	-0,379
9.	0,150	-0,456
10.	0,141	-0,666
11.	-0,077	-1,931
12.	-0,166	0,247
13.	0,091	0,673
14.	0,118	2,590
15.	0,001	-1,615
16.	-0,014	-0,673
17.	-0,048	-0,268
18.	0,119	0,214

Berdasarkan hasil analisis pada subtes 2 (kemampuan mengenali asumsi) diketahui bahwa hanya terdapat dua butir yang memiliki daya beda > 0,3, yaitu butir nomor 3 dan 4. Hal ini menunjukkan bahwa kedua butir tersebut yang dapat dengan cukup baik membedakan probabilitas menjawab benar antara individu yang berkemampuan dalam mengenali asumsi dengan individu yang berkemampuan tinggi. ICC dari butir nomor 3 dan 4 yang disajikan pada Gambar 2.

Dari 18 butir dalam subtes 2, terdapat 8 butir yang memiliki daya beda bertanda negatif. Hal tersebut mengindikasikan terjadi kesalahan dalam penulisan butir yang membuat individu yang berkemampuan tinggi dalam mengenali asumsi, gagal me-

mahami maksud dari butir tersebut. Butir tersulit terdapat pada butir no 2 dengan tingkat kesulitan 2,448 artinya hanya individu dengan kemampuan minimal sebesar 2,448 yang memiliki 50% peluang untuk menjawab butir dengan benar. Butir yang termudah adalah butir no 4 dengan nilai tingkat kesulitan -3,920, artinya peserta dengan kemampuan dalam mengenali asumsi sebesar -3,920 berpeluang 50% menjawab benar butir tersebut.



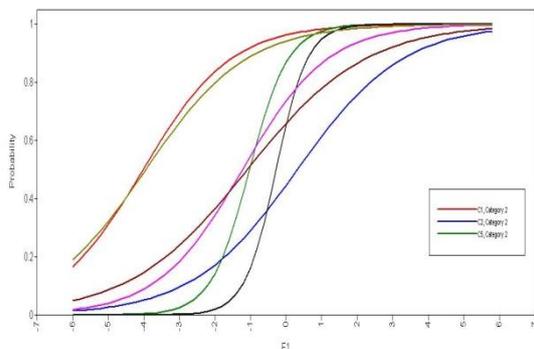
Gambar 2. ICC Subtes 2

Tabel 4. Hasil Estimasi Parameter Butir Subtes 3

No. Butir	Daya Beda (a)	Tingkat Kesulitan (b)
1.	0,477	-3,254
2.	0,398	0,220
3.	-0,276	0,479
4.	-0,376	2,164
5.	1,086	-1,888
6.	1,340	-0,630
7.	0,493	-1,023
8.	0,355	-0,656
9.	0,414	-2,777
10.	0,038	-2,063

Berdasarkan hasil analisis terhadap subtes 3 (deduksi: kemampuan menilai kelogisan sebuah kesimpulan) diketahui bahwa terdapat 7 butir yang memiliki nilai daya beda lebih besar dari 0,3. Hal tersebut menunjukkan bahwa sebagian besar butir pada subtes 3 dapat membedakan antara individu dengan kemampuan deduksi yang tinggi dengan yang rendah. Ketujuh butir tersebut memiliki bentuk ICC dengan kemiringan yang cenderung curam. Hal ini me-

nunjukkan bahwa butir tersebut memiliki daya beda yang cukup baik sehingga dapat menggambarkan individu yang berkemampuan tinggi memiliki peluang menjawab benar yang lebih besar daripada yang berkemampuan rendah. ICC dari butir nomor 1, 2, 5, 6, 7, 8, 9 disajikan pada Gambar 3.



Gambar 3. ICC Subtes 3

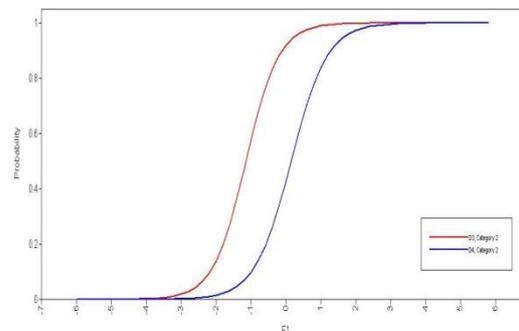
Pada subtes 3 terdapat dua butir yang memiliki daya beda bertanda negatif, yaitu butir nomor 3 dan 4. Hal ini mengindikasikan terjadinya kesalahan dalam penulisan butir, sehingga gagal dipahami oleh individu dengan kemampuan yang tinggi. Butir nomor 4 merupakan butir dengan tingkat kesulitan paling tinggi sedangkan butir yang termudah adalah butir nomor 1.

Tabel 5. Hasil Estimasi Parameter Butir Subtes 4

No. Butir	Daya Beda (a)	Tingkat Kesulitan (b)
1.	0,101	1,761
2.	0,258	0,583
3.	1,256	-2,435
4.	1,139	0,290
5.	-0,120	0,179
6.	-0,111	0,344
7.	-0,176	0,049
8.	0,013	-1,581
9.	-0,051	1,162
10.	-0,089	-1,177

Pada subtes 4 (interpretasi: kemampuan menilai sebuah *evidence*) hanya terdapat dua butir yang memiliki daya beda yang cukup baik ($> 0,3$) yaitu butir nomor 3 dan 4. Artinya, kedua butir ini dapat membeda-

kan antara individu yang berkemampuan tinggi dengan yang berkemampuan rendah didalam kemampuan menilai apakah bukti yang digunakan dalam penarikan kesimpulan telah sesuai dengan data yang tersedia. ICC untuk butir nomor 3 dan 4 disajikan pada Gambar 4.



Gambar 4. ICC Subtes 4

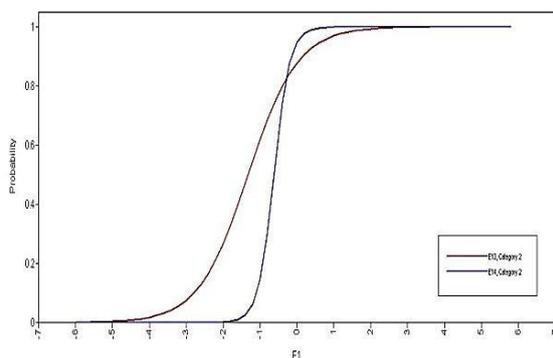
Terdapat 5 butir yang memiliki daya beda dengan nilai negatif. Kelima butir ini tergolong butir yang buruk sehingga individu dengan kemampuan yang tinggi gagal memahami maksud dari butir tersebut. Butir yang tersulit terdapat pada butir nomor 1 sedangkan yang termudah terdapat pada butir nomor 3.

Tabel 6. Hasil Estimasi Parameter Butir Subtes 5

No. Butir	Daya Beda (a)	Tingkat Kesulitan (b)
1.	-0,058	1,037
2.	-0,103	-0,453
3.	0,056	-0,752
4.	-0,069	0,221
5.	0,135	-1,188
6.	0,234	-0,666
7.	0,058	-0,251
8.	-0,011	-0,079
9.	0,220	0,018
10.	-0,043	-0,318
11.	0,143	-0,274
12.	0,120	-0,096
13.	0,880	-1,975
14.	2,764	-2,904
15.	-0,495	-0,544

Pada subtes 5 (kemampuan mengevaluasi argumen) hanya terdapat dua butir

yang memiliki daya beda yang tinggi, yaitu butir nomor 13 dan 14. Kedua butir ini mampu dengan baik membedakan individu yang berkemampuan tinggi dalam menilai sebuah argumen dengan individu yang berkemampuan rendah. Dengan demikian, ICC pada kedua butir ini akan memiliki kecuraman tinggi sehingga menunjukkan perbedaan bahwa individu yang berkemampuan tinggi akan memiliki peluang menjawab benar yang lebih besar daripada individu yang berkemampuan rendah. ICC dari butir nomor 13 dan 14 disajikan pada Gambar 5.



Gambar 5. ICC Subtes 5

Terdapat 6 butir yang memiliki daya beda negatif, artinya terjadi kesalahan penulisan yang membuat kesalahan informasi sehingga individu yang berkemampuan tinggi gagal memahami butir. Kesulitan tertinggi terdapat pada butir nomor 1 dan butir yang termudah adalah butir nomor 14. Revisi terhadap butir-butir dengan nilai daya beda yang rendah maupun negatif akan dilakukan pada tahap pengembangan tes berikutnya.

Berdasarkan hasil analisis dari kelima subtes dalam tes kemampuan berpikir kritis, diperoleh data bahwa dari total 68 butir hanya terdapat 15 butir yang memiliki daya beda lebih besar dari 0.3. Hal ini menunjukkan bahwa ke 15 butir tersebut dapat membedakan probabilitas menjawab benar antara individu yang berkemampuan tinggi dengan yang berkemampuan rendah.

Dari keseluruhan butir terdapat 27 butir dengan daya beda yang negatif. Hal ini menunjukkan bahwa banyak terjadi kesalahan dalam penulisan butir sehingga terjadi kesalahan dalam memahami informasi pada

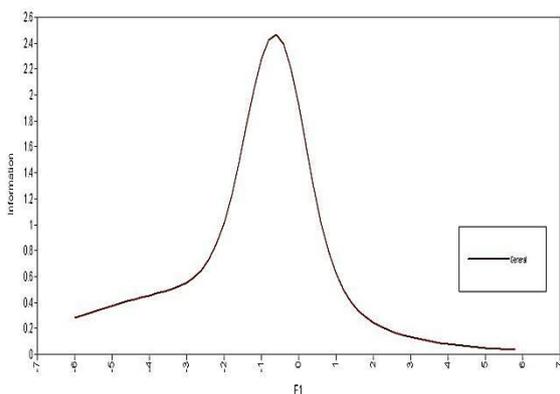
individu yang memiliki kemampuan berpikir kritis tinggi. Banyaknya butir yang berdaya beda rendah serta bertanda negatif menunjukkan bahwa terdapat banyak kesalahan dalam proses penulisan butir. Hal ini dapat terjadi karena pemilihan topik yang digunakan sebagai skenario pada batang soal kemungkinan besar bukanlah topik yang dapat digunakan untuk mengukur kemampuan berpikir kritis.

Pemilihan topik pada penelitian ini telah dilakukan oleh para ahli yaitu para dosen sekaligus Psikolog di bidang Industri dan Organisasi serta Psikolog di bidang Pendidikan dan Perkembangan. Pemilihan pakar tersebut mungkin saja kurang tepat karena belum tentu menguasai bidang berpikir kritis dengan baik. Artinya, terjadi kesalahan dalam menentukan *Subject Matter Experts* di bidang berpikir kritis (Possin, 2014).

Kesalahan lainnya dimungkinkan terjadi dalam menentukan jawaban benar. Soal-soal yang terdapat pada tes kemampuan berpikir kritis berbeda dengan soal-soal yang terdapat pada tes objektif misalnya di bidang matematika, ilmu alam maupun pengetahuan lain yang jawaban benar dari soal tersebut telah diketahui dengan pasti. Soal pada tes yang sejenis kemampuan berpikir kritis memiliki cara skoring yang berbeda. Pada jenis tes yang tidak memiliki cara skoring baku terdapat tiga pilihan cara skoring, yaitu (a) *target-scoring*, (b) *expert-scoring*, (c) *consensus-scoring* (MacCann, 2006). Dalam penelitian ini digunakan cara skoring dengan metode *expert-scoring*. Kondisi ini memungkinkan terjadinya kesalahan karena prosedur kesalahan dalam menentukan *Subject Matter Experts* di bidang berpikir kritis. Kesalahan-kesalahan ini akan diperbaiki pada proses penelitian selanjutnya yaitu pada tahap penelitian tahun kedua.

Seperti yang tampak pada gambar 6, dapat diketahui bahwa nilai fungsi informasi Tes Berpikir Kritis akan maksimum pada tingkat kemampuan $-0,5$ yaitu sebesar 2,4. Berdasarkan grafik di atas juga dapat diketahui bahwa tes Berpikir kritis yang dikembangkan dapat mengestimasi dengan

tepat tingkat kemampuan yang berada pada rentang -3 sampai dengan 1.



Gambar 6. Test Information Function

Simpulan

Meskipun belum optimal dan masih terdapat sejumlah catatan yang perlu diperbaiki pada penelitian berikutnya, namun penelitian ini telah menghasilkan tes berpikir kritis yang dapat digunakan dalam proses diagnosis maupun seleksi dalam lingkungan akademik dan kerja. Terkait proses, setiap tahapan dalam pengembangan tes juga telah dilakukan sesuai dengan yang dikemukakan oleh Hambleton dan Jones (1993).

Penelitian lanjutan di tahun kedua sangat diperlukan untuk memperbaiki hasil yang telah diperoleh dengan memperhatikan berbagai catatan yang telah diberikan pada bagian pembahasan

Daftar Pustaka

- Angelo, T.A. & Cross, P. (1995). *Classroom assessment techniques: a Handbook for college teachers* (2nd edition). San Francisco: Jossey-Bass Publishers.
- Baker, F. B. (2001). *The basics of butir response Theory* (2nd edition). USA: ERIC Clearinghouse on Assessment and Evaluation.
- Embretson, S. E., & Reise, S. P. (2000). *Butir response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of butir*

response theory. Newbury Park, CA: Sage Publications.

- Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on: Comparison of classical test theory and butir response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3). 38-47.
- Hossoubah, Z. (2007). *Developing creative and critical thinking skills (terjemahan)*. Bandung: Yayasan Nuansa Cendia.
- Kaupp, J., Frank, B., & Chen, A. (2014). *Evaluating critical thinking and problem solving in large classes: Model eliciting activities for critical thinking development*. Toronto: Higher Education Quality Council of Ontario.
- Kurfiss, J. G. (1988). *Critical thinking: Theory, research, practice, and possibilities*. (ASHE-ERIC Higher Education Report No. 2). Association for the Study of Higher Education. <http://www.eric.ed.gov/PDFS/ED304041.pdf> (diunduh 14 April 2014).
- Lai, E. R. (2011). *Critical thinking: a Literature review*. Author, Pearson Assessments. <http://www.pearsonassessments.com/hai/images/tmrs/CriticalThinking> (diunduh 14 April 2014).
- MacCann, C.E. (2006). New approaches to measuring emotional intelligence: Exploring methodological issues with two new assessment tools. *Doctoral dissertation*. Sydney: University of Sydney.
- Norris, S. P. & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
- Possin, K. (2014). Critique of the Watson-Glaser critical thinking appraisal test: The more you know, the lower your score. *Informal Logic*, Vol 34, No 4.
- Stassen, M.L.A., Herrington, A., & Henderson, L. (2011). *To improve the academy*. San Francisco, CA: Jossey-Bass Publishers.

- Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement*. National Institute of Education. <http://eric.ed.gov/PDFS/ED272882.pdf> (diunduh 14 April 2014).
- Essential Skills for the 21st Century Workplace: Keys to Succeeding in the Global Economy. www.21stcenturyskills.org (diunduh 14 April 2014)
- Wagner, T. A., (2002). Critical thinking: The development of a new measure. *Thesis*. Blacksburg, Virginia: Virginia Polytechnic Institute and State University.
- Walker, P. & Finney, N. (1999). *Skill development and critical thinking in higher education*. London: Higher Education Research & Development Unit, University College.