

UJI RELIABILITAS INSTRUMEN KEMAMPUAN BERPIKIR KRITIS SCTT  
(*SURABAYA CRITICAL THINKING TEST*)

Oleh:

Chikitta Grace Parera

Fakultas Psikologi, Universitas Surabaya

**Abstrak** - Kemampuan berpikir kritis sangat penting adanya. Pentingnya kemampuan berpikir kritis dapat terlihat ketika seseorang menghadapi masalah-masalah yang menuntutnya untuk melakukan *problem solving*, membuat keputusan serta meregulasi diri terhadap masalah-masalah sosial di lingkungan sekitarnya. Untuk mengetahui kualitas kemampuan berpikir kritis seseorang tersebut, sangat diperlukan sebuah alat ukur yang biasa disebut instrumen. Instrumen yang baik haruslah diuji reliabilitas dan validitasnya. Instrumen kemampuan berpikir kritis di Indonesia adalah instrumen *surabaya critical thinking test* yang dibuat pada tahun 2017.

Penelitian ini dilakukan guna menguji reliabilitas instrumen *surabaya critical thinking test*. Uji reliabilitas yang dilakukan adalah uji reliabilitas antar penilai. Penelitian ini melibatkan 34 partisipan. Uji reliabilitas dilakukan dengan menggunakan Cohen's Kappa.

Hasil yang didapatkan dalam penelitian ini ialah berkisar pada range 0.434 sampai 0.946 dan nilai sig yang dihasilkan adalah sama yaitu 0.000. Dari nilai tersebut, diketahui bahwa nilai Kappa instrumen *surabaya critical thinking test* di atas 0.400 sehingga dapat dikatakan reliabel. Namun, terdapat satu item pada instrumen *surabaya critical thinking test* yang menunjukkan bahwa terjadi kesepakatan antar penilai yang cenderung lebih rendah daripada item lainnya yakni *surabaya critical thinking C* item 2B yang mempunyai nilai Kappa terendah yaitu 0.434. Hal tersebut disebabkan oleh adanya perbedaan persepsi antar penilai 1 dan 2 terhadap jawaban subjek.

***Kata kunci - Cohen's kappa, Kemampuan berpikir kritis, Surabaya Critical Thinking, Uji reliabilitas.***

***Abstract*** – The capacity to think critical is very important. The importance of the capacity to think critical can be seen when someone face the problems which is demand people to do problem solvin, make decision and regulate himself against social problem in the their environmen. To know the quality of ability critical thinking, we can measuring with something that we called as instrument. An good instrument shall tested a reliability and validaty. One of an instrument that measure critical thinking in Indonesia is Surabaya Critical Thinking Test. which made on 2017

This research is done to test reliability an instrument Surabaya Critical Thinking Test. Reliability test excuted is reliability test between assessors. This research involving 34 participants. Reliability test was conducted using Cohen's Kappa.

The result obtained in this research is on the range 0.434 until 0.946 and value sig is generated the same is 0.000. From these values, noted that Kappa value of the Instrument Surabaya Critical Thinking Test above 0.400 so it can be said to be good reliability. However, there is one item on the instrument Surabaya Critical Thinking Test showed that there is agreement between assessors the to be lower than the other items namely Surabaya Critical Thinking Test C item 2B which has lowest kappa value is 0.434. It is caused by the difference in perception between the assessors 1 and 2.

***keywords- Cohen's kappa, Critical Thinking, Surabaya Critical Thinking, Reliability Test***

## **PENDAHULUAN**

Kemampuan berpikir kritis adalah salah satu *soft skill* yang sangat dibutuhkan keberadaannya bagi seseorang untuk menghadapi permasalahan dalam

kehidupan sehari-hari. Pada aplikasi nyata kehidupan sehari-hari, kemampuan berpikir kritis sangatlah penting. Hal tersebut dapat dilihat ketika kemampuan berpikir kritis digunakan pada saat seseorang akan mengambil sebuah keputusan, melakukan *problem solving* serta meregulasi diri. Ketika seseorang menerima sebuah informasi baru, kemampuan berpikir kritis pun digunakan untuk mengevaluasi dan menganalisis informasi tersebut, sehingga tidak akan ada berita *hoax* atau berita yang bertujuan membuat provokasi yang menyebar di masyarakat.

Kualitas kemampuan berpikir kritis tersebut dapat diketahui dengan adanya pengukuran kemampuan berpikir kritis secara khusus. Kemampuan berpikir kritis dapat diukur dengan menggunakan alat ukur atau yang sering disebut sebagai instrumen.

Dewasa ini di Indonesia cukup banyak dilakukan penelitian tentang kemampuan berpikir kritis. Melihat hal tersebut, para ahli mulai melakukan penelitian mengenai sebuah instrumen secara khusus yang mengukur kemampuan berpikir kritis. Salah satu contoh yang dilakukan oleh Siti Zubaidah (2017), ia melakukan sebuah pengembangan instrumen kemampuan berpikir kritis dengan format esai. Dalam pengembangannya, ia belum secara jelas mencatumkan perihal uji validitas dan reliabilitasnya, sehingga dapat disimpulkan bahwa instrumen tersebut belum memenuhi kaidah.

Melihat fenomena tersebut, terdapat dorongan yang mencetuskan seorang ahli untuk membuat skala pengukuran atau instrumen kemampuan berpikir kritis yang lebih baik adanya sehingga terciptalah *surabaya critical thinking test*.

Seperti yang dikatakan oleh Azwar (2015) yang mengatakan sebuah tes psikologi yang baik memiliki tiga hal yang diperhatikan antara lain : a) jenis data skor yang diperoleh oleh tes, b) reliabilitas data hasil tes, dan c) masalah validitas data yang dihasilkan oleh tes.

Dalam pengembangan *surabaya critical thinking test*, instrumen kemampuan berpikir kritis yang baru adanya. Belum ada satupun uji kaidah-kaidah psikometri yang dilakukan. Hal tersebut mendorong bahwa penting adanya uji kaidah-kaidah psikometri seperti salah satunya ialah uji reliabilitas. Menguji reliabilitas instrumen *surabaya critical thinking test* adalah salah satu upaya menguji validasi konsistensi internal dari instrumen tersebut guna pengembangan instrumen lebih lanjut.

*Surabaya critical thinking test* adalah instrumen kemampuan berpikir kritis yang dibuat secara khusus dalam bahasa indonesia. Instrumen yang merupakan golongan jenis tes potensi ini dibuat oleh Anindito Aditomo pada tahun 2017. *Surabaya critical thinking test* terdiri atas 3 model yakni *surabaya critical thinking test A, B, dan C*. Pada setiap instrumen terdapat soal permasalahan yang berbeda-beda serta jumlah butir yang sama yakni 3 *multiple choice* dan 4 esai.

Menurut Reynold & Livingston (2012) mengatakan bahwa ada beberapa jenis uji reliabilitas seperti uji reliabilitas tes ulang, uji reliabilitas tes paralel, uji reliabilitas konsistensi internal dan antar penilai. Dari setiap jenis uji reliabilitas memiliki kelemahan masing-masing. Uji reliabilitas tes ulang contohnya, tes ini memiliki kelemahan *carryover effect* bila dilakukan dalam interval waktu yang terlalu dekat sehingga dibutuhkan interval waktu yang cenderung panjang untuk

mendapatkan hasil yang baik (Reynold & Livingston, 2012). Selain itu, uji reliabilitas paralel memiliki kelemahan pada sulitnya membuat soal format paralel dan cenderung membutuhkan waktu yang banyak dalam proses membuat soal tersebut (Reynold & Livingston, 2012). Dari kelemahan-kelemahan tersebut, dapat disimpulkan bahwa akan membutuhkan waktu yang cukup panjang untuk menguji reliabilitas dengan metode uji reliabilitas tes ulang dan paralel.

Kelemahan uji reliabilitas antar penilai terletak pada subjektivitas penilaian. Namun, kelemahan dari kedua metode uji reliabilitas yaitu uji reliabilitas tes ulang dan paralel, akan dapat diminimalisasi dengan metode uji reliabilitas antar penilai. Hal inilah mendorong peneliti untuk memilih uji reliabilitas antar penilai.

Berdasarkan latar belakang tersebut sehingga melakukan penelitian yaitu tentang uji reliabilitas instrumen kemampuan berpikir kritis Surabaya Critical Thinking Test.

#### **METODE PENELITIAN**

Penelitian ini menggunakan metode penelitian *research & development*. Penelitian *research & development* dilakukan dengan melakukan proses validasi, yaitu penelitian yang digunakan guna pengembangan sebuah alat ukur atau sebuah instrumen (Arifin, 2016). Validasi atau proses validitas memiliki arti sebagai ukuran ketepatan, kecermatan dan keakuratan sebuah alat ukur melakukan fungsinya (Siaputra & Natalya, 2016). Reliabilitas adalah salah satu unsur dari validitas struktur internal. Reliabilitas dinilai penting dalam proses validasi sebab reliabilitas akan mengukur utilitas sebuah instrumen (Siaputra & Natalya, 2016).

## HASIL DAN PEMBAHASAN

Uji reliabilitas antarpemilai dilakukan dengan metode Cohen's Kappa menggunakan *IBM SPSS Statistic 22*. Persentase yang terjadi antara pemilai 1 dan pemilai 2 dijabarkan sebagai berikut.

*Persentase kesepakatan antara pemilai instrumen SCTT A, SCTT B dan SCTT C*

Instrumen	Persentase (%)	Kappa	T	$\rho$ value
SCTT A no 1B	91.2	0.727	5.189	0.000
SCTT A no 2A	88.2	0.784	5.252	0.000
SCTT A no 2B	76.5	0.626	4.998	0.000
SCTT A no 3C	97.1	0.946	8.584	0.000
SCTT B no 1B	82.4	0.690	4.595	0.000
SCTT B no 2A	82.4	0.707	5.484	0.000
SCTT B no 2B	82.4	0.713	5.880	0.000
SCTT B no 3C	82.4	0.695	5.665	0.000
SCTT C no 1B	94.1	0.767	4.600	0.000
SCTT C no 2A	82.4	0.682	5.174	0.000
SCTT C no 2B	61.8	0.434	3.913	0.000
SCTT C no 3C	85.3	0.631	5.684	0.000

Berdasarkan tabel di atas, dapat dilihat nilai uji reliabilitas Cohen's Kappa di item 1B,2A,2B dan 3C pada instrumen SCTT A, B, dan C berkisar pada nilai terendah yaitu 0.434 dan di nilai tertinggi yaitu 0.946. Di sisi lain, jika melihat nilai kesepakatan antar pemilai di item 1B,2A,2B dan 3C pada instrumen SCTT A,B dan C yaitu berkisar pada nilai terendah yaitu 61.8% dan yang tertinggi yaitu 97.1 %. Sedangkan nilai  $\rho$  value adalah sama di item 1B, 2A, 2B dan 3C pada instrumen SCTT A, B dan C yaitu 0.000.

Dengan melihat nilai uji reliabilitas Cohen's Kappa, presentase kesepakatan, dan  $\rho$  value dapat disimpulkan bahwa item 1B, 2A, 2B dan 3C di

instrumen SCTT A,B dan C ialah reliabel. Namun, terdapat 1 item yaitu item 2B pada instrumen SCTT C yang memiliki nilai Kappa yang cukup rendah yaitu 0.434. Nilai Kappa tersebut dikatakan reliabel dengan nilai yang rendah dengan mengetahui bahwa batas nilai reliabilitas adalah di atas 0.400.

Hal-hal yang mengakibatkan nilai reliabilitas Kappa item 2B instrumen *Surabaya Critical Thinking Test* sangat rendah jika dibandingkan dengan semua item yang diujikan nilai reliabilitasnya yaitu karena adanya perbedaan persepsi perbedaan persepsi antara penilai 1 dengan penilai 2 tentang pemetaan jawaban subjek dalam tiga kelompok yaitu:

1. Jawaban subjek yang tidak menyanggah argumen berbeda dari orang lain maupun mendukung argumen subjek sebelumnya
2. Jawaban subjek hanya mendukung argumen sebelumnya namun tidak menyanggah argumen berbeda dari orang lain.
3. Jawaban subjek mendukung jawaban subjek sebelumnya dan mampu menyanggah argumen berbeda dari orang lain.

Adanya perbedaan persepsi antara antara penilai 1 dan 2 ini sebagian besar diakibatkan oleh jawaban subjek yang bersifat ambigu atau bermakna implisit.

#### **KESIMPULAN DAN SARAN**

Terdapat 1 item yaitu item 2B instrumen *Surabaya Critical Thinking Test C* yang memiliki nilai reliabilitas cukup rendah jika dibandingkan dengan seluruh item yang diujikan yaitu 0.434.

Hal-hal yang mengakibatkan nilai reliabilitas Kappa item 2B instrumen *Surabaya Critical Thinking Test* sangat rendah yaitu karena adanya perbedaan persepsi perbedaan presepsi antara penilai 1 dengan penilai 2 yang diakibatkan oleh jawaban subjek yang bersifat ambigu atau bermakna implisit.

Nilai *output* kappa intrumen dapat juga diakibatkan kurang maksimalnya pengembangan panduan skoring *Surabaya Critical Thinking Test* yang digunakan untuk melakukan skoring sehingga memperoleh nilai reliabilitas tersebut. Ketika melihat panduan skoring *Surabaya Critical Thinking Test*, terdapat 1 fitur yang belum dicantumkan secara jelas yaitu fitur kriteria evaluasi.

Dalam pengembangan panduan skoring *Surabaya Critical Thinking Test*, bila terjadi ketidaksepakatan antar kedua penilai maka penilai melakukan hal yakni melakukan komunikasi dua arah perihal evaluasi pada masing-masing penilai serta telah dilakukannya pelatihan pada penilai kedua sebelum dilakukannya skoring.

Guna menghasilkan hasil penelitian lanjutan yang lebih maksimal, ada beberapa hal yang dapat dijadikan langkah-langkah untuk ditempuh. 1) Penelitian selanjutnya akan lebih maksimal bila dapat mencakup subjek dari kedelapan fakultas yang ada di Universitas Surabaya. 2) Penelitian akan lebih maksimal bila jumlah subjek yang diambil untuk penelitian ialah sama untuk setiap fakultasnya. 3) melakukan uji validasi dengan metode lainnya seperti uji reliabilitas tes ulang atau uji validitas. 4) melakukan perubahan kasus pada *Surabaya Critical Thinking Test C* dengan kasus yang lebih *complicated* sehingga menimbulkan banyak presepsi dari subjek. 5) Setelah perbaikan dilakukan, instrumen dapat



diaplikasikan secara praktis dalam pengujian kemampuan berpikir kritis dalam kancan mahasiswa, perguruan tinggi dan masyarakat.

## DAFTAR PUSTAKA

- Arifin, Z. (2016). Pengembangan instrumen pengukur kemampuan berpikir kritis matematika siswa SMA kelas X. *THEOREMS (The original research of Mathematics)*, 1: 1, 58-74.
- Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Common misconceptions of critical thinking. *J. Curriculum studies*, 31, 269-283.
- Baron, J. (2008). *Thinking and deciding fourth edition*. United States: Cambridge University Press.
- Ennis, R. H. (1985). A logical basic for measuring critical thinking skills. *Educational Leadership*, 45-48.
- Ennis, R. H. (1991). Critical Thinking : A Streamlined Conception. *Teaching Philosophy*, 14:1, 5-25.
- Ennis, R. H. (1993). Critical Thinking Assessment. *Theory into practice*, 179-186.
- Facione, P. A. (1990). *The California Critical Tinking Skills Test College Leve: Technical report #1 Experimental validation nd content validity*. California: California Academic Press.
- Facione, P. A., & Facione, N. C. (2009). *The holistic critical thinking scoring rubric: A tool for developing and evaluating critical thinking*. California: California Academic Press.
- Kilem Li Gwet, P. (2014). *Handbook of Inter-Rater Reliability: The definitive guide to measuring the extent of agreement among raters / by Kilem Li Gwet, PhD- 4th ediion*. United States: 3-24.
- Lai, E. R. (2011). *Critical Thinking : a review literature*. Pearson Assessment .
- Mandrekar, J. N. (2011). Measures of interrater agreement. *Journal of thoracic oncology*, 6:1, 6-7.
- PhD, K. I. (2014). *Handbook of inter0rater reliability fourth edition : the definitive guide to measuring the extent of agreement among raters*. united states: Advanced Analytics, LLC.
- Reddy, Y. ..., & Andrade, H. (2009). A review of rubric use in higher education. *Assessment & evaluation in higher education*, 35:1, 435-448.

Reynolds, C. R., & Livingston, R. B. (2011). *Mastering modern psychological testing : theory & methods*. California: Pearson .

Sternberg, R. J. (1986). *Critical Thinking : its nature, measurement and improvement*. Washington DC: National Inst. of education .

Viera, A. J., & Joanne M. Garrett, P. (2005). Understanding interobserver agreement : The Kappa Statistic. *Family Medicine*, 37: 5, 360-363.

Willingham, D. T. (2007). Critical thinking : why it is so hard to teach? *American Educator*, 8-18.

