BAB III. METODE PENELITIAN

3.1. Lokasi, Subyek dan Waktu Penelitian

Penelitian ini dilakukan di Universitas Negeri Malang. Populasi penelitian adalah seluruh mahasiswa yang menempuh matakuliah Pendidikan Kewarganegaraan atau Pendidikan Pancasila pada semester I 2015/2016 yang tersebar di Universitas Negeri Malang. Sampel penelitian adalah 184 mahasiswa yang menempuh matakuliah Pendidikan Kewarganegaraan atau Pendidikan Pancasila pada semester I 2015/2016 yang tersebar di seluruh fakultas di Universitas Negeri Malang yang terpilih secara random. Mahasiswa dikelompokkan dalam kelompok sains dan sosial. Yang tergolong sains adalah fakultas matematika dan IPA, fakultas teknik, sedangkan yang tergolong sosial adalah fakultas ilmu sosial, fakultas ekonomi, fakultas bahasa, dan fakultas olah raga. Penentuan subyek penelitian sebagai kelompok perlakuan dan pembandung dilakukan dengan teknik random yang dilakukan oleh koordinator perkuliahan Pendidikan Pancasila dengan memperhatikan karakteristik sains dan sosial.

Waktu penelitian adalah paruh kedua semester genap tahun kuliah 2015/2016, yang mulai dilaksanakan 6 Oktober sampai 25 Desember 2015.

3.2. Pendekatan Penelitian

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan kuantitatif. Metode penelitian yang dipakai adalah *Quasi-experimental* (Kuasi - eksperimen). Mengapa dipilih pendekatan kuantitatif? Karena penelitian ini menekankan pada pengujian untuk melihat perbandingan rata-rata dua kelompok yang berbeda yang dilakukan terhadap data yang digunakan berdasarkan alat ukur kuantitatif dan menggunakan dan alat analisis yang kuantitatif.

Alasan lain dipilihnya pendekatan kuantitatif adalah sisi efisiensi. Analisis kuantitatif bekerja menggunakan sampel untuk memecahkan persoalan yang dihadapi. Selain dari sisi sample, metode kuantitatif ini dapat memberikan penjelasan yang lebih tepat terhadap fakta kuantitatif yang diperoleh.

Menurut pengelompokan eksperimen dari Stanley (1967) sebenarnya penelitian ini tergolong dalam dalam *True Experimental* karena menggunakan

kelompok kontrol dengan melakukan pre dan post tes. Namun penelitian ini lebih tepat menggunakan penyebutan *Quasi-experiment* (Eksperimen Semu atau kuasi-eksperimen) karena subyek penelitian adalah manusia yang tentunya tidak dapat diperlakukan sebagai obyek eksperimen yang murni.

Desain penelitian ini menggunakan *pre-post test control-experiment group design.* Ciri utama rancangan ini adalah : (1) penempatan kelompok eksperimen dan kontrol dipilih secara acak, (2) kelompok eksperimen diberi perlakuan Model Analisis Nilai Berbantuan Internet, sedangkan kelompok kontrol diberikan perlakuan pengajaran konvensional. Jadi yang membedakan keduanya adalah pembelajarannya, sedangkan karakteristik lain sama. Pembelajaran model analisis nilai berbantuan internet menggunakan langkah-langkah pembelajaran sesuai dengan langkah pembelajaran model analisis nilai dari Metcalf dan Jerrold Coombs, sedangkan pembelajaran konsvensional adalah pembelajaran sebagaimana yang selama ini dilakukan di kelas perguruan tinggi yang melibatkan penggunaan metode diskusi, penugasan, tanya jawab dan ceramah. (3) sebelum diberi perlakuan pembelajaran mahasiswa diberi *pretest* dan sesudah selesai perlakuan diberikan *Post Test.*

3.3. Desain Penelitian

Desain penelitian ini meliputi dua kelompok mahasiswa di Universitas Negeri Malang Kelas eksperimen yang dipilih secara acak dan kelas kontrol sebagai pembanding. Setiap mahasiswa mengikuti pre tes dan pos tes, CCTST dan CCTDI, untuk mengumpulkan data yang relevan dengan pertanyaan riset dari penelitian ini.

Desain Penelitian

Bentuk desain penelitian yang digunakan dalam penelitian ini adalah eksperimen dengan *pre-and-post test control group design.* Desain eksperimen dapat dilihat pada gambar

R O_1	**x**	**O_2**
R O_3	**-**	**O_4**

Gambar 2.4. Desain Penelitian

Desain Eksperimen Semu dengan *Pre-post test control group design.*

Keterangan:

R = Random, yang menempatkan kelas sebagai kelompok perlakuan atau pembanding

O_1= Observasi satu yaitu kelompok eksperimen yang mendapat pre test

X = Perlakuan Model Analisis Nilai Berbantuan Internet

O_2= Observasi dua yaitu kelompok eksperimen yang mendapat post test

O_3= Observasi tiga yaitu kelompok kontrol yang mendapat pre test

- = Tanpa perlakuan Model Analisis Nilai Berbantuan Internet*/pembelajaran konvensional*

O_4= Observasi empat yaitu kelompok kontrol yang mendapat post test.

Desain ini dipilih karena dapat lebih cermat membedakan perubahan itu dengan membandingkan dengan kelompok, baik yang sebelum maupun sesudah perlakuan.

Berikut ini disajikan penjabaran dari masing-masing variabel yang digunakan dalam penelitian ini.

No	Variabel	Sub Variabel	Deskriptor	Riset Aktu
1	Model Analisis Nilai Berbantuan Internet	6 Langkah Pembelajaran dan 4 Pengujian	Diuraikan di bahasan selanjutnya	Penerapan langkah-langkah pembelajaran
2	Keterampilan Berpikir Kritis	Analisis Interpretasi Inferensi Evaluasi Penjelasan induksi deduksi	Diuraikan di bahasan selanjutnya	CCTST
3	Disposisi Berpikir Kritis	Pencarian Kebenaran Keterbukaan Pikiran Analitis Sistematis Kepercayaan pada Penalaran Rasa Ingin Tahu Kematangan Pertimbangan	Diuraikan dibahasan selanjutnya	CCTDI

3.4. Validitas dan Reliabilitas CCTST

CCTST telah menjalani sejumlah tes psychometric oleh para pengembang. Berikut ini disajikan hasil uji validitas dan reliabilitas yang telah dilakukan pengembang. Informasi tentang validitas dan reliabilitas ini diperoleh dari berbagai manual CCTST yang dikirimkan oleh lembaga penerbit untuk saling melengkapi (Manual CCTST 2015 dan *Insight Assessment*, 2016, Manual CCTST 2013).

Dalam hal item pilihan ganda sebagai metode untuk memperoleh bukti yang sah dari kompetensi seperti berpikir kritis, ilmu pengukuran memberikan bukti jelas bahwa keterampilan kognitif tingkat tinggi, seperti berpikir kritis, dapat diukur secara sah dan andal dengan item pilihan ganda yang ahli *(well-crafted multiple choice items).* Dimulai tahun 1970-an Program penelitian CCTST, yang ditunjukkan melalui metode analisis item psikometrik biasa dan satu set analisis protokol yang dikembangkan oleh Peter dan Noreen Facione bahwa pilihan item berbingkai tertutup dapat membedakan pemikir yang kuat dari pemikir lemah dalam sampel peserta tes lintas disiplin ilmu dan bidang profesional. Validasi tambahan Dr. Joanne Carter Wells dari California State University Fullerton dan Dr. Carol Gittens dari Santa Clara University yang merupakan kolaborator dalam pekerjaan ini untuk membentuk tes keterampilan berpikir dan perlengkapannya serta mendapat sumbangan banyak konsultan ahli profesi-spesifik dari sektor bisnis swasta, militer, kesehatan, pendidikan dan profesi hukum.

3.4.1. Validitas Konten/Isi *(Content Validity)*

Validitas Konten/isi mengacu pada kemampuan tes untuk mendapatkan alat ukur dari domain yang dimaksud. Identifikasi domain yang bersangkutan, dan mendapatkan kecocokan atasnya, merupakan kepentingan utama untuk validasi isi (Cronbach, 1990). Kriteria kedua validitas isi adalah meyakinkan bahwa metode konstruksi tes yang "masuk akal" dilakukan. Dalam kasus CCTST tersebut, domain khususnya adalah berpikir kritis seperti yang didefinisikan dan dibahas dalam Bab 2. Berpikir kritis, seperti yang didefinisikan oleh studi APA Delphi, adalah suatu konstruk yang mengintegrasikan sejumlah kesatuan kognitif yang dikenal sebagai komponen dari jenis proses penalaran manusia ini, termasuk

dalam laporan studi Delphi APA sebagai konsep yang menyertakan, misalnya, Analisis, Inferensi. dan Evaluasi. CCTST ini dirancang sebagai alat ukur holistik terhadap konstruk Berpikir Kritis, dengan mencantumkan skala yang dapat digunakan untuk menguji konsep yang tercakup.

Validitas Konten/Isi CCTST ini lebih didukung oleh pilihan yang dibuat oleh pendidik dalam bidang penalaran, peneliti, dan akademisi disertasi doktor yang mempelajari keterampilan penalaran manusia, serta profesional sumber daya manusia yang ingin mempekerjakan karyawan dengan keterampilan keputusan yang kuat, yang mengadopsi CCTST tersebut. Validitas pengukuran juga mensyaratkan bahwa instrumen pengujian harus bebas dari pengganggu yang tidak diinginkan *(unintended distractors)* yang mempengaruhi pilihan jawaban kelompok peserta tes dan dicocokkan untuk kelompok peserta tes yang diinginkan. Dalam penelitian ini menggunakan CCTST 2000 yang memang diperuntukkan oraang dewasa dan mahasiswa.

Dalam tes keterampilan berpikir kritis California, peserta tes ditantang untuk membentuk pertimbangan rasional yang didasarkan skenario singkat yang disajikan dalam pertanyaan. CCTST tidak menguji pengetahuan konten. Pertanyaan CCTST dibingkai dalam konteks kegiatan sehari-hari. Semua informasi penting yang diperlukan untuk menjawab pertanyaan dengan benar disajikan dalam pertanyaan. Fakta bahwa CCTST hanya mengukur berpikir kritis dan bukan pengetahuan konten memungkinkan untuk menggunakan instrumen ini sebagai pretest dan posttest untuk mengukur peningkatan berpikir kritis yang terjadi selama program pendidikan dalam hal ini perkuliahan Pendidikan Pancasila. Untuk alat ukur yang valid tentang berpikir kritis, instrumen harus menyajikan berbagai kesulitan yang sesuai bagi individu atau kelompok yang diuji yang memungkinkan penskalaan skor yang akurat. Tes keterampilan berpikir kritis CCTST dirancang untuk menyertakan form yang benar tentang CCTST untuk menguji kekuatan dan kelemahan dalam berpikir kritis dalam berbagai kisaran menyeluruh tentang individu atau kelompok.

Validitas konten dilakukan pada tahap pengembangan CCTST sebelumnya, melibatkan 46 ahli nasional yang ikut dalam survei Delphi. Validitas konkuren *(Concurrent validity)* diukur dengan menghubungkan

CCTST dengan skor masuk perguruan tinggi seperti skor *Scholastic Aptitude Test Verbal* (*correlation alpha* = 0.55) and dengan GRE (= 0.72).

3.4.2. Validitas Konstruk (*Construct Validity*)

Validitas konstruk ditunjukkan oleh studi korelasional yang mengkorelasikan skor berpikir kritis dengan alat ukur lain yang pokok isinya mencakup konstruk. Bentuk-bentuk CCTST telah menunjukkan korelasi yang kuat dengan instrumen lain yang mempunyai isi yang mencakup alat ukur pemikiran kritis atau penalaran tingkat tinggi sebagai komponen skor atau peringkatnya. Telah terbukti ada korelasi tinggi dengan tes standar kesiapan tingkat perguruan tinggi dalam penalaran tingkat tinggi (Skor Total GRE: r Pearson = 0,719, p <.001; GRE *Analytic* r = 0,708, p <.001; GRE *verbal* r = 0,716, p <.001; GRE kuantitatif, r = 0,582, p <.001). Sejumlah hubungan ini dilaporkan dalam studi penelitian multi-situs besar yang melibatkan 50 program pendidikan ilmu kesehatan yang menilai keterampilan berpikir mahasiswa.

Validitas konstruk didukung oleh korelasi signifikan antara CCTST dan gender, etnis, akademik utama dan kepercayaan diri mahasiswa pada saat diadministrasi untuk 1196 skor mahasiswa universitas (Facione, 2006).

Secara teoritis, nilai rata-rata (IPK)/*Grade Point Average (GPA)* seharusnya berkorelasi dengan skor kemampuan berpikir kritis. Sebagian laporan temuan menunjukkan hubungan ini kecil atau tidak ada signifikansi (r = .20, p <.oo1), mungkin fungsi dari sedikit variasi dalam IPK di berbagai sampel atau kenyataan bahwa proporsi dari varians yang dijelaskan dalam keterampilan berpikir hanya kecil/sedikit yang dimiliki oleh alat ukur IPK holistik. Usia bukan prediktor signifikan dari kemampuan berpikir kritis ketika tingkat pendidikan dikontrol; ini merupakan temuan yang konsisten dalam sampel ukuran yang memadai.

Bukti terhadap validitas konstruk dari CCTST ditunjukkan oleh adanya peningkatan skor tes CCTST mahasiswa setelah mereka mengambil matakuliah berpikir kritis atau pelatihan program pendidikan berpikir kritis penalaran klinis. Beberapa publikasi tinjauan sejawat pilihan ditulis oleh peneliti di seluruh dunia sebagai contoh bagaimana berbagai versi CCTST telah digunakan untuk dokumentasi tambahan dalam keterampilan berpikir kritis. Peningkatan skor

harus disebabkan perbaikan dalam berpikir kritis dan bukan karena berbagai faktor eksternal lainnya. Dengan kata lain, semua variabel tetap konstan dengan satu pengecualian yaitu treatmen yang diberikan yang diharapkan dapat meningkatkan keterampilan berpikir kritis. Ini mungkin, misalnya, program pendidikan (dalam hal ini mata kuliah Pendidikan Pancasila) difokuskan pada analisis berbasis kasus tentang masalah nyata (dalam hal ini model analisis nilai) ketika penekanannya adalah pada pelatihan keterampilan berpikir kritis, suatu model pembelajaran berpikir kritis yang melatih mahasiswa yang bekerja dalam penggunaan keterampilan berpikir kritisnya. Kemudian, akan masuk akal mengharapkan bahwa skor posttest lebih baik untuk individu atau kelompok yang sama dapat dikaitkan pada efek dari intervensi model untuk membangun keterampilan dan disposisi berpikir kritis. Untuk memaksimalkan kualitas dalam kondisi pengujian, peneliti berkonsultasi dengan staf teknis dari *Insight Asesment* dalam desain rencana pengujian yang dibuat (yang bernama Crist Smith).

Dalam penelitian awal, tes keterampilan berpikir kritis California telah dilaporkan mendapatkan keuntungan keterampilan berpikir kritis setelah hanya beberapa minggu pelatihan keterampilan berpikir kritis. Beberapa program pemberian gelar profesional telah menunjukkan hasil yang signifikan dalam keterampilan berpikir kritis dengan menggunakan kurikulum spesifik dalam disiplin professional. Studi validasi juga telah dilakukan untuk menunjukkan bahwa tidak ada bukti ancaman terhadap validitas karena pengalaman sebelumnya dengan instrumen tes (uji efek). Dalam studi pembahasan yang ketat, ketika CCTST digunakan pada posttest dua minggu setelah pretest, peserta tes diamati untuk mengingat beberapa butir item melalui pilihan jawaban lagi untuk menentukan respon mereka terhadap setiap item. Hasil dalam skor tes CCTST seharusnya diharapkan ketika intervensi pelatihan berpikir kritis dirancang dan dijalankan dengan baik dan ketika peserta tes termotivasi secara memadai untuk mengikuti tes pada posttest. Umpan balik dari administrator tes baik dalam latar pendidikan maupun tempat kerja menunjukkan bahwa peserta tes menemukan tes CCTST yang sulit tapi menarik.

3.4.3. Validitas Kriteria *(Criterion Validity)*

Validitas Kriteria adalah pertimbangan yang paling penting dalam validitas tes. Validitas kriteria mengacu pada kemampuan tes untuk memprediksi suatu kriteria perilaku eksternal pada tes itu sendiri. *Criterion validity refers to the ability of the test to predict some criterion behavior external to the test itself* (Nunnally, JC dan Bernstein, IH., 1994) Misalnya, seseorang mungkin ingin tahu bahwa CCTST dapat memprediksi beberapa pengukuran yang bermakna untuk menunjukkan prestasi hasil belajar yang ditunjukkan atau penyiapan kesuksesan dan lisensi profesional kunci dalam masyarakat, atau transisi yang berhasil ke tempat kerja. Hasil yang ditunjukkan skor CCTST memberikan jenis nilai prediktif pada penelitian review sejawat yang diterbitkan secara independen. (Beberapa di antaranya : Williams, Glasnapp, Tilliss, Osborn, Wilkins, Mitchell, Kershbaum, Schmidt (2003), Giddens dan Gloeckner (2005), Denial (2008); Vendrely (dalam *Insight Assessment, 2016*)

Dalam studi validasi CCTST awal, analisis dilakukan untuk meneliti apakah mahasiswa yang menyelesaikan matakuliah selama satu semester yang memerlukan berpikir kritis menunjukkan peningkatan dalam keterampilan berpikir kritis dibandingkan dengan mahasiswa yang tidak menempuh matakuliah tersebut. Matakuliah berpikir kritis dalam penelitian ini telah disetujui oleh komite universitas yang mengawasi persyaratan pendidikan berpikir kritis. Penelitian ini, yang melakukan treatmen dan desain kelompok kontrol, menggunakan baik *cross-sectional* maupun alat ukur pretest-posttest berpasangan *(matched-pairs).* Peningkatan yang signifikan terlihat baik di cross-sectional ($t = 2,44$, satu ekor $p < 0,008$) maupun analisis-berpasangan ($t = 6.60$, $df = 231$, $p < .001$). Rata-rata mahasiswa dalam sampel berpasangan berubah dari persentil ke-55 ke persentil ke-70 (persentil posttest berdasarkan pada kelompok pretest). Kelompok kontrol (tanpa intervensi pendidikan berpikir kritis khusus) baik dalam eksperimen *cross-sectional* maupun berpasangan tidak menunjukkan hasil yang signifikan.

3.4.4. Reliabilitas *(Reliability)*

Instrumen pengujian *Insight Assessment* telah memenuhi ambang batas untuk reliabilitas konsistensi internal yang kuat (Alpha minimal 0,80 untuk alat ukur atribut dan minimum KR-20 dari 0,72 untuk alat ukur keterampilan) dan

diamati untuk mempertahankan kinerja ini di semua sampel varians yang memadai.

Studi validasi dilakukan dalam pengembangan CCTST tentang tes keterampilan dan alat ukur atribut yang terkait untuk memastikan bahwa topik item tetap menarik saat ini dan relevan dengan budaya kelompok peserta tes.

Data dari studi validasi berkelanjutan menghasilkan estimasi konsistensi reliabilitas internal Estimasi KR-20 (Kuder Richardson - 20) terkini untuk bentuk *CCTST* tingkat perguruan tinggi adalah 0,78 – 0,82. Kuder-Richardson -20 adalah statistik sebanding dengan Alpha Cronbach yang digunakan untuk instrumen dan skala yang diskor secara dikotomi. Untuk instrumen dengan skala multidimensi, KR-20 di atas 0,70 menunjukkan tingkat tinggi dari konsistensi internal.

Koefisien ini menunjukkan bukti yang lebih dari cukup untuk menginspirasi kepercayaan pada konsistensi internal instrumen ini dalam mengukur konstruk keseluruhan: berpikir kritis. Skala memiliki item yang lebih sedikit dari instrumen secara keseluruhan, namun reliabilitas konsistensi internalnya tetap kuat. KR - 20 untuk skala secara konsisten berkisar 0,52 hingga 0,77. Beban faktor untuk item pada masing-masing sub skala berkisar 0,300 hingga 0,770.

Studi reliabilitas tambahan dari terjemahan resmi dari CCTST telah dilakukan. Tes terjemahan (baru) dikeluarkan jika koefisien reliabilitas tercapai pada sampel yang memadai dari penutur bahasa asli yang diuji dengan CCTST terjemahan. Sekalipun koefisien reliabilitas 0.80 disarankan untuk konsistensi internal pada instrumen yang dimaksudkan untuk menargetkan tunggal, kemampuan homogen, koefisien lebih rendah sering terlihat dalam kasus tes yang diskor secara dikotonomi, terutama tes yang menguji konsep multifaset seperti berpikir kritis. Peringkat reliabilitas 0,65 hingga 0,75 dipertimbangkan mencukupi untuk menempatkan kepercayaan dalam instrumen jenis ini (Nunnally, 1994).

3.5. Validitas dan Reliabilitas CCTDI.

Berikut ini akan disajikan validitas dan reliabilitas instrumen CCTDI yang digunakan dalam penelitian ini.

3.5.1. Validitas Konten *(Content Validity)*

Kriteria pertama untuk validitas konten merujuk pada kemampuan penilaian alat ukur itu untuk menangkap ranah yang dikehendaki. Identifikasi ranah yang terkait, dan diperolehnya kesepakatan sangat penting untuk validasi konten. Disusun berdasarkan Deskripsi Pemikir Ideal dari Delphi APA, CCTDI dikembangkan untuk menangkap atribut yang dimasukkan di dalam deskripsi aspek disposisi berpikir kritis. Validitas konten dari item CCTDI ini berasal dari data Delphi APA asli, dan diperluaskan melalui review pendidik untuk semua tingkatan.

Kriteria kedua dari validitas konten adalah jaminan telah dilakukannya cara pembuatan instrumen yang layak *("sensible").* CCTDI merupakan alat ukur sikap dan dikonstruksi dengan menggunakan cara khas untuk jenis pengukuran ini. Konfirmasi CCTDI mengekspresikan pendapat umum, keyakinan, nilai-nilai, harapan dan persepsi, dan tidak menggunakan kosa kata teknis atau jargon pemikiran kritis. Berbagai konfirmasi item percontohan ditulis untuk menangkap deskripsi konsensus tentang pemikir kritis ideal. Pengembangan konfirmasi item dalam bentuk item sikap memberi ciri pada pengembangan CCTDI. 250 item yang dihasilkan disaring oleh pendidik berpikir kritis tingkat perguruan tinggi dan peneliti yang ahli dalam penelitian survey dan pengembangan instrumen mengidentifikasi item tersebut untuk menghilangkan ambiguitas dan salah penafsiran.

Ketika instrumen diterjemahkan ke bahasa lain, cara mempertahankan validitas item dan skala merupakan hal yang tak terpisahkan. Variasi budaya mengharuskan beberapa item itu diubah dalam terjemahan non bahasa Inggris karena isu interpretasi budaya dan idiomatik, maka berbagai versi bahasa tidak sama persis pada tingkat item. Walau demikian, berbagai kehati-hatian diambil, melalui kolaborasi dengan pakar internasional yang merupakan pemilik bahasa asli, dan dengan menggunakan prosedur penerjemahan ketat berulang-ulang untuk memastikan validitas, reliabilitas dan kompetensi budaya agar dicapai dalam terjemahan resmi. Analisis berulang dari instrumen selanjutnya dilakukan dalam sampel bahasa Inggris dan dalam terjemahan.

CCTDI telah dikutip dalam sejumlah besar literatur mencerminkan temuan di seluruh dunia. Hasil penelitian terdahulu menyertakan penelitian disertasi

doktor yang menguji pemikiran kritis dalam hubungan dengan pelatihan disiplin dalam berbagai bidang ilmu. Demikian juga penelitian independen yang meneliti atribut pola pikir dari berpikir kritis dalam hubungan dengan pencapaian hasil belajar dengan menggunakan berbagai strategi pendidikan seperti pembelajaran berbasis-masalah, jurnal, dan peta-pikiran, dan penggunaan skenario berbasis kasus dalam seminar dan diskusi pelatihan. Studi riset dilakukan peneliti dari berbagai negara di seluruh dunia, dengan menggunakan instrumen CCTDI dalam berbagai bahasa. Misalnya, dengan menggunakan metode korelasi Canon, Kwon, Onwuegbuzie dan Alexander melaporkan hubungan antara skor disposisi berpikir kritis dan ketertarikan yang berkaitan dengan penggunaan pustaka akademik. Keingintahuan, Sistematis dan Kepercayaan adalah prediktor signifikan dari ketertarikan pustaka, dengan kepercayaan menjelaskan persentase terbesar dari varian. Dalam studi Swedia, Wangensteen dkk. (2011) meneliti disposisi berpikir kritis sebagai prediktor penelitian yang menggunakan praktek klinik.

3.5.2. Validitas Permukaan *(Face Validity)*

Validitas Permukaan *("Face Validity")* adalah istilah yang sering digunakan orang untuk merujuk kesan seseorang saat membaca item tes yang diberikan dan mengambilnya pada nilai permukaan. Orang memandang instrumen sebagai memiliki validitas permukaan yang baik jika item yang muncul menunjukkan domain yang relevan. Pada tes pengetahuan faktual, validitas permukaannya jelas. Namun, pada alat ukur atribut kepribadian, validitas permukaan dapat menjadi masalah. Risikonya adalah bahwa peserta tes tidak menjawab dengan sebenarnya karena ia berharap yang dirasakan merupakan sesuatu yang positif. Ini disebut "Bias Respons Keinginan Sosial." Jika survey sebelum kerja meminta pelamar kerja untuk setuju atau tidak setuju dengan "Saya seorang pekerja yang dapat diandalkan." Seseorang yang ingin dianggap sebagai karyawan berpotensi yang diharapkan akan menjawab "setuju" bahkan jika jawaban yang sesungguhnya adalah "tidak setuju." Cara memerangi risiko ini adalah menggunakan item yang kurang transparan atau tidak mudah dilihat langsung. Misalnya, daripada menanyai pelamar kerja itu setuju atau tidak setuju dengan sejumlah pernyataan, survey itu menanyakan beberapa hal positif dan negatif, yang mengungkap kepercayaan, nilai dan niat mengenai perilaku

karyawan yang dapat diandalkan. Misalnya "Karena perjalanan panjang buat pulang pergi cukup beralasan bila terlambat datang kerja beberapa menit pagi-pagi." CCTDI dirancang dengan menggunakan pendekatan ini untuk mengatasi ancaman bias respons keinginan sosial.

Bias respon keinginan sosial tidak diketemukan pada CCTDI *(Manual CCTDI 2016). Marlowe-Crowne Social Desirability Scale* dipakai untuk mengevaluasi resiko bias respon keinginan sosial pada alat ukur laporan diri tertentu. Sebenarnya Marlowe-Crowne adalah alat ukur keadaan sebenarnya dari jawaban orang pembuatnya. Jika orang menjawab sesuai dengan respon yang diinginkan benar secara sosial, akan ada korelasi tinggi skor antara Marlowe-Crowne dan skor pada instrumen yang dievaluasi.

Dalam kasus alat ukur CCTDI, tidak ada kaitan signifikan teramati antara skor skala CCTDI dan skor Marlowe-Crowne. Item CCTDI telah mampu membedakan dengan baik antar responden, yang selalu mendukung item dari berbagai atribut pola berpikir kritis. Misalnya, orang yang rendah skor Pencarian Kebenaran umumnya setuju dengan item yang mengindikasikan penentangan terhadap pencarian kebenaran. Mereka gagal melihat bahwa menyetujui pernyataan demikian menghasilkan deskripsi diri yaitu respon tidak diinginkan secara sosial dan tidak menyenangkan. Begitu juga pemikir tidak teratur. Mereka cenderung menyetujui item yang mendeskripsikan item yang tersebar dan tidak sistematis dalam pendekatannya terhadap pemecahan masalah dan pengambilan keputusan. Dengan kata lain, evaluasi Marlowe-Crowne dari CCTDI mendukung instrumen karena menunjukkan bahwa peserta tes menjawab secara jujur, dan bukannya mengutamakan jawaban yang diinginkan.

3.5.3. Reliabilitas *(Reliability)*

Instrumen asesmen CCTDI telah memenuhi ambang batas reliabilitas konsistensi internal yang kuat (minimal Alfa 0,80 untuk alat ukur disposisi) dan diobservasi mempertahankan performa ini pada semua sampel dari varian yang memadai. Reliabilitas konsistensi internal untuk tujuh skala masing-masing di dalam sampel percontohan CCTDI awal berkisar dari 0.71 untuk 0.80, dengan alfa untuk instrumen keseluruhan mencapai 0.91. Nilai-nilai yang kuat telah diobservasi secara konsisten dalam sampel yang dikumpulkan selama lebih dari

15 tahun (mulai dari 0.60 hingga 0.78 pada skala dan 0.90 atau ke atas untuk alat ukur keseluruhan). Koefisien Reliabilitas yang lebih rendah teramati pada sampel kecil di mana varian skornya tidak besar. Studi Internal Reliabilitas test ulang menghasilkan koefisien CCTDI yaitu melebihi 0.80 ketika dua waktu administrasi diberikan dua minggu setelah waktu pertama pada sampel di mana tidak ada program pendidikan yang berlangsung.

3.6. Prosedur Penelitian

Prosedur penelitian dilakukan dengan melalui tahapan sebagai berikut:

3.6.1 Pra Pengembangan

Prosedur penelitian tahap pertama adalah melakukan kajian pustaka, jurnal, laporan penelitian, disertasi, internet, serta diskusi dengan ahli dan dosen untuk mengkaji tentang Model Analisis Nilai Berbantuan Internet untuk mengenali landasan filsofis, teori yang mendasari dan langkah-langkah pembelajarannya. Selanjutnya mencari berbagai sumber yang berkaitan dengan instrumen penelitian.

3.6.2. Pengembangan Instrumen

Pengembangan instrumen tidak dilakukan oleh penulis, namun oleh lembaga *Insight Assessment* yang mengelola instrumen baku keterampilan dan disposisi berpikir kritis CCTST dan CCTDI.

3.6.2.1. Instrumen Penelitian Keterampilan Berpikir Kritis

Instrumen penelitian mencakup instrumen pengukuran keterampilan (CCTST) dan disposisi berpikir kritis (CCTDI) dari *Insight Assessment. California Critical Thinking Skills Test (CCTST)* 2000 digunakan untuk mengumpulkan data tentang kemampuan dan keterampilan berpikir kritis partisipan. CCTST adalah instrumen pilihan ganda 34 item terstandar yang menantang intelektual yang diberikan selama periode 45 menit untuk melakukan analisis, interpretasi, inferensi dan evaluasi melalui penggunaan *diagrammatic and text-based contexts.* Item didasarkan pada topik yang netral disiplin ilmu, tanpa peranan jenis kelamin dan konteks stereotype kelas sosial, jangka pendek, pernyataan masalah; dan skenario dikelompokkan ke dalam enam skor dari CCTST: keseluruhan, dan lima skor sub skala yang meliputi analisis, evaluasi, inferensi, induksi dan deduksi (*Insight Assessment 2016,*

2015, 2013, Facione & Facione, 2002, hlm. 15). Penskoran disimpulkan komputer dengan CAPSCORE, layanan dari California Academic Press. Menurut Facione (2006), kemungkinan skor pada total skor CCTST dapat terentang dari 0 hingga 34. Masing-masing dari 34 items yang nampak pada CCTST dialokasikan pada salah satu dari tiga pertama kelompok sub skala (Analisis, Evaluasi, dan Inferensi) dan tiga puluh item dari tiga puluh empat itu tersebar pada item terakhir (Induksi dan deduksi) yang didaftar di bawah ini.

Tabel 3.1. Kisi-kisi Skala, Sub Skala, dan Sebaran Item Tes CCTST

Sub-skala	Butir Tes	Keterangan
Analisis	1 – 9	Interpretasi dan analisis
Evaluasi	10 – 13	Evaluasi
	25-34	Penjelasan
Inferensi	14 – 24	Inferensi
Induksi	30 di antara 1 – 34	Tiga puluh dari tiga puluh empat item secara simultan diklasifikasi induktif dan deduktif
Deduksi		

Disarikan dari Facione (1990d)

Tabel 3.2. Rekomendasi Asesmen Kinerja Skor Total CCTST

REKOMENDASI ASESMEN KINERJA Skor Keseluruhan	Skor Keseluruhan CCTST – Rekomendasi Asesmen Kinerja				
	Tidak Nampak	Lemah	Sedang	Kuat	Superior
Skor Keseluruhan CCTST Versi 2000 34-poin	0 – 7	8 - 12	13 – 38	19 – 23	24 atau lebih tinggi

Deskripsi Skor Keseluruhan Rekomendas Asesmen Kinerja

Superior/Unggul: Hasil ini menunjukkan keterampilan berpikir kritis yang lebih unggul dibandingkan sebagian besar peserta tes. Keterampilan di tingkat superior konsisten dengan potensi untuk belajar dan kepemimpinan yang lebih maju.

Kuat: Hasil ini konsisten dengan potensi keberhasilan akademis dan karir.

Sedang: Hasil ini menunjukkan potensi tantangan yang berhubungan dengan keterampilan ketika terlibat dalam pemecahan masalah reflektif dan pengambilan keputusan reflektif terkait dengan pembelajaran atau pengembangan karyawan.

Lemah: Hasil ini prediksi mengalami kesulitan dengan tuntutan yang terkait dengan pendidikan dan lapangan kerja untuk pemecahan masalah reflektif dan pengambilan keputusan reflektif.

Tidak Nampak: Hasil ini konsisten dengan kemungkinan upaya yang tidak memadai peserta tes, kelelahan kognitif, atau mungkin kesulitan membaca atau pemahaman bahasa.

Tabel 3.3. Rekomendasi Asesmen Kinerja Skor Skala CCTST

REKOMENDASI ASESMEN KINERJA Skor Skala CCTST	Rekomendasi Asesmen Kinerja Skor Skala CCTST 2000 (versi 34-poin)		
	Tidak Nampak	**Sedang**	**Kuat**
Analisis	0 - 2	3 – 4	5 atau lebih
Inferensi	0 - 5	6 – 11	12 atau lebih
Evaluasi	0 - 3	4 – 7	8 atau lebih
Induksi	0 - 5	6 – 11	12 atau lebih
Deduksi	0 - 5	6 – 11	12 atau lebih

(Diadaptasi dari *CCTST User Manual, 2016)*

Lebih jauh lagi, sub-skala skor pada instrumen dapat berkisar sebagai berikut: Analisis (0 sampai 9); Inferensi (0 sampai 14); evaluasi (0 sampai 11); penalaran deduktif (0 hingga 16); dan penalaran Induktif (0 hingga 14). Jumlah skor dari analisis, evaluasi, dan membuat kesimpulan adalah sama dengan total nilai CCTST. Skala induktif dan deductif tumpang tindih dengan skala analisa, membuat kesimpulan, dan evaluasi (Facione, 2006).

3.6.2.2. Instrumen Penelitian Disposisi Berpikir Kritis

Instrumen penelitian untuk mengukur disposisi berpikir kritis adalah California Critical Thinking Disposition Inventory (CCTDI). CCTDI yang membutuhkan 15 – 30 menit untuk menyelesaikan, tidak berisi kosa kata teknis berpikir kritis dan diasumsikan tidak ada kandungan materi tingkat perguruan tinggi. 75 item mendorong pengambil tes untuk "express familiar opinions, beliefs, values, expectations, and perceptions" dengan menggunakan enam-poin yang diminta melakukan pilihan skala Likert berkisar dari “sangat setuju” hingga “sangat tidak setuju” (Facione, Facione, & Giancarlo, 2001, hlm. 6).

Berikut ini disajikan kisi-kisi skala CCTDI yang terdapat pada instrumen.

Tabel 3.4. Kisi-Kisi Skala, Sub Skala, dan Sebaran Item Tes CCTDI

Sub Skala Disposisi Berpikir Kritis	Aspek yang dinilai	Sebaran Item tes
Pencarian Kebenaran	disposisi terhadap pencarian	Tujuh subskala atau

Sub Skala Disposisi Berpikir Kritis	Aspek yang dinilai	Sebaran Item tes
	kebenaran atau bias,	atribut dari CCTDI tersebar secara acak pada 75 item
Keterbukaan	disposisi terhadap keterbukaan pikiran atau intoleransi,	
Analitis	mengantisipasi kemungkinan konsekuensi atau melalaikannya,	
Sistematis	mengerjakan dengan cara yang sistematis atau tidak sistematis,	
Percaya pada penalaran	menuju menjadi percaya diri dalam kekuatan penalaran atau ketidak percayaan dalam berpikir,	
Keingintahuan	ke arah yang ingin tahu atau bertahan terhadap pembelajaran,	
Kematangan Keputusan	Ke arah pertimbangan matang dan bernuansa atau ke arah pemikiran sederhana yang kaku.	

Tabel 3.5. Ringkasan Analisis Interpretasi Skor CCTDI Menurut *Insight Assessment*

Rentangan Skor Numerik CCTDI	Kategori Kualitatif CCTDI	Interpretasi Atribut CCTDI yang Direkomendasikan
Skor 50 - 60	Positif Kuat	Skor ini mengindikasikan atribut kebiasaan pikiran yang positif dan kemungkinan pada faktor ke pendekatan individu untuk semua pemikiran urutan lebih tinggi (definisi masalah reflektif dan pemecahan masalah reflektif), khususnya ketika situasi konsekuensi tinggi
Skor 40 - 49	Positif	Skor ini mengindikasikan penguasaan dan tetap konsisten dari atribut yang diukur dan umumnya menunjukkan atribut pola pikir yang dapat dipercaya/andal
Skor 30 - 39	Tidak konsisten/ ambivalen	Skor ini mengindikasikan tidak tegas/tidak konsisten dari sikap yang diukur. Individu sering terlihat bergerak dari kategori skor ini menuju rentangan lebih tinggi sebagai hasil program pelatihan atau pendidikan untuk mengembangkan atribut pola pikir ini. Jarang, upaya tersebut menyebabkan individu turun ke rentangan yang lebih merugikan atau berlawanan pada asesmen kedua. Pergerakan terhadap keengganan/ ketidaksukaan atau pertentangan ini konsisten dengan observasi lain dari formasi sikap. Kebiasaan berpikir dalam kehidupan dan kerja yang menuntut keterampilan berpikir urutan lebih tinggi memerlukan keberanian, ketekunan, kejujuran, dan organisasi. Mungkin ada beberapa individu menolak tantangan tersebut atau terlalu takut atau kesulitan.
Skor 20 – 29	Negatif	Skor ini mengindikasikan penilaian lemah atau enggan terhadap atribut yang diukur, kebiasaan pikiran negatif, dan mendekati penalaran dan pemecahan masalah

Rentangan Skor Numerik CCTDI	Kategori Kualitatif CCTDI	Interpretasi Atribut CCTDI yang Direkomendasikan
		dengan keterbatasan kualitas berpikir dan pembuatan keputusan, dan kurang berpikiran terbuka dan kurang kemauan terlibat masalah dalam beberapa situasi. Jarang terjadi pergerakan dari tingkat skor ini ke level kuat terlihat pada beberapa individu setelah partisipasi dalam kesempatan pendidikan, lebih umum terlihat individu mengekspresikan pada ambivalen pada alat ukur saat asesmen kedua
Skor 10 – 19	Negatif Kuat	Skor ini adalah indikasi kenegatifan atau penentangan kuat terhadap atribut yang diukur. Skor ini menunjukkan bahwa individu mungkun secara sistematis menghindari situasi yang memerlukan pemikiran dan pengambilan keputusan atau bekerja mengganggu atau merusak proses pengambilan keputusan yang orang lain usahakan membuatnya. Pergerakan dari rentangan skor ini ke rentangan skor positif setelah partisipasi dalam sebuah kesempatan pendidikan untuk memperkuat akribut jarang teramati.

Untuk konteks penelitian ini, peneliti tidak perlu lagi melakukan uji coba kelayakan dan validasi bahasa karena sudah dilakukan lembaga penerbit instrumen yang menggunakan tenaga ahli Indonesia Asli dan sudah terikat ketentuan tidak melakukan perubahan apapun pada instrumen dan diharuskan mengirimkan kembali instrumen untuk dimusnahkan setelah digunakan. Penulis membeli instrumen dengan status hak pakai saja.

Tes ini sudah dalam bentuk bahasa Indonesia dan telah diuji coba oleh tim yang terdiri dari pakar orang Indonesia. CCTST diperlukan maksimal 45 menit masa tes, sedangkan CCTDI dalam waktu 30 menit. Dengan ketentuan bahwa jika mahasiswa mengerjakan kurang dari 10 menit atau mengerjakan 60 persen maka data itu tidak akan dihitung karena dianggap sebagai data palsu.

Dalam konteks penelitian disertasi ini, CCTST dan CCTDI diberikan secara tulis bukan online karena pertimbangan bahasa. Peneliti sudah menghubungi *Insight Assessment* untuk mengkaji kemungkinan memakai sistem online agar hasil bisa langsung diketahui dan mempercepat pengolahan lanjutan demi kepentingan disertasi. Namun kendalanya terletak pada bahasa. Uji online hanya tersedia dalam bahasa Inggris dan setelah konsultasi dengan mahasiswa disepakati untuk menggunakan tes tulis yang menggunakan bahasa Indonesia.

Dalam studi ini dua kelompok besar (yang terbagi dalam 3 kelas kelompok eksperimen dan 3 kelas kelompok kontrol) yang terdiri dari 184 mahasiswa. Penggunaan sampel yang cukup besar untuk eksperimen ini merupakan upaya peneliti untuk mencegah klaim yang menyesatkan *(misleading)* dari hasil analisis tentang pengaruh Model Analisis Nilai Berbantuan Internet pada keterampilan dan disposisi berpikir kritis. Untuk masing-masing kelompok (baik dalam pre maupun pos tes), kinerja mahasiswa rata-rata pada masing-masing tes dihitung dan dikumpulkan berdasarkan kelompok. Perbedaan hasil tes antara kelompok perlakuan dan kelompok pembanding dianalisis untuk menentukan signifikansi statistik pada unit masing-masing dengan uji t dan anova.

Perangkat pertanyaan riset ketiga dan seterusnya digolongkan pada karakteristik jenis kelamin, asal jurusan di SMA, dan asal daerah yang mungkin berkontribusi pada penjelasan perbedaan signifikan dalam pengaruh Model Analisis Nilai Berbantuan Internet pada mahasiswa, yang dinyatakan dengan analisis data terkait. Mahasiswa, bukan kelompok, adalah unit analisis yang cocok untuk langkah penelitian ini.

3.6.3. Pengembangan bahan dan Sumber Belajar

Bahan perkuliahan sudah ditentukan oleh tim universitas dan penelitian ini hanya mengubah strategi pembelajaran pada paruh kedua perkuliahan dengan model analisis nilai berbantuan internet dan pengajaran konvensional.

3.6.4. Evaluasi

Kegiatan evaluasi untuk kepentingan penelitian hanya pada tes CCTST dan CCTDI, sedangkan evaluasi materi untuk kepentingan nilai perkuliahan tidak termasuk dan tidak disertakan dalam laporan penelitian ini karena tidak langsung terkait dengan analisis data guna menjawab pertanyaan penelitian.

Uji coba terdiri dari tiga tahap, yaitu menguji langkah-langkah pembelajaran Model Analisis Nilai Berbantuan Internet melalui konsultasi dengan dosen pengajar mata kuliah Pendidikan Pancasila, uji coba model pembelajaran dan uji coba menjawab soal CCTST dan CCTDI.

(1) Uji Coba Pertama

Uji coba awal pada tim uji ahli untuk mengkaji ketepatan langkah-langkah pembelajaran Model Analisis Nilai Berbantuan Internet dengan melibatkan tiga

dosen (satu orang bergelar S3 dan dua bergelar S2 yang mengajar matakuliah Pendidikan Pancasila sebagai justifikasi ahli untuk mengkaji dasar-dasar teoritis akademis).

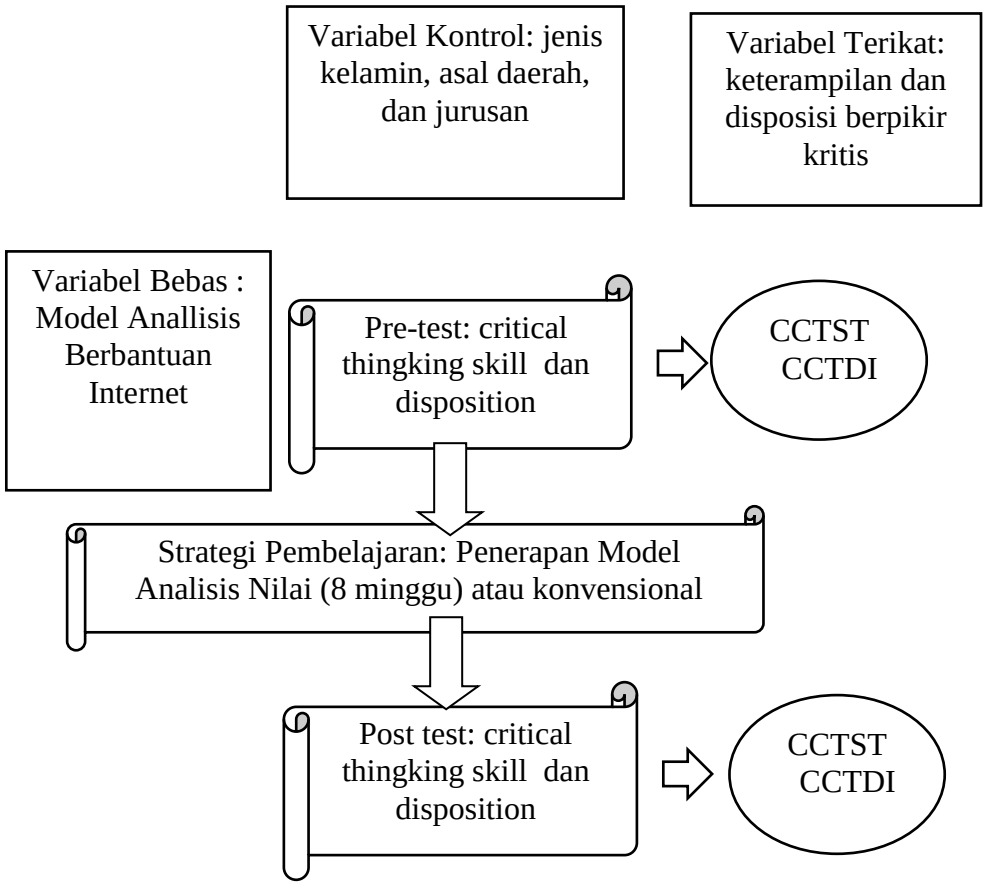
(2) Uji Coba Kedua

Uji coba lebih ditujukan pada langkah-langkah pembelajaran model analisis nilai dan bukan pada variabel dosen Pendidikan Pancasila. Uji coba langkah-langkah pembelajaran dilakukan peneliti pada semester sebelumnya pada kelas yang lain.

(3) Uji coba Ketiga

Uji coba ketiga diberikan pada mahasiswa untuk menjawab 3 (tiga) soal yang diunduh secara online dengan menggunakan soal yang setara namun berbeda dengan instrument penelitian. Sampel instrumen (tiga instrumen) juga sudah dikonsultasikan dengan promotor dan ko-promotor. Sampel ditetapkan dan diambil dari enam kelas.

3.6.5. Pelaksanaan



Gambar 2.3. Pelaksanaan Perlakuan

Studi ini membandingkan pengajaran Model Analisis Nilai Berbantuan Internet dan pembelajaran konvensional dalam hal keterampilan dan disposisi berpikir kritis.

3.6.6. Peserta dan Pengaturan

Peneliti menggunakan sampel dari 184 peserta. Sampel memasukkan semua mahasiswa peserta yang menempuh mata kuliah pendidikan Pancasila tahun pertama di Universitas Negeri Malang yang menggunakan Model Analisis Nilai Berbantuan Internet (N = 92) dan mahasiswa yang mendapat perlakuan pembelajaran berbasis kuliah tradisional (N = 92). Kedua kelompok berasal dari populasi mahasiswa Universitas Negeri Malang yang memiliki karakteristik yang sama. Mahasiswa dari kedua kelompok tersebut mendapat konten yang sama karena dikelola oleh koordinator MDKU yang menjamin konsistensi pengajar MDKU sebagai matakuliah universitas. Hanya perbedaan di antara dua kelompok adalah strategi pembelajaran. Mahasiswa Model Analisis Nilai Berbantuan Internet menggunakan langkah-langkah sesuai sintaknya. Sementara, mahasiswa kelompok pembanding menggunakan perkuliahan yang biasanya dilakukan oleh dosen melalui pemberian metode diskusi, tanya jawab, ceramah dan pemberian tugas.

3.6.7. Intervensi Strategi Pembelajaran

Silabus dan konten matakuliah untuk kedua kelompok dirancang oleh koordinator akademis yang sama yang mengkoordinasikan semua matakuliah yang diajarkan di MKDU tingkat universitas. Kedua kelompok (kontrol dan eksperimen) menerima materi pembelajaran, buku teks, penugasan dan evaluasi sama, namun kedua kelompok menggunakan metode pengajaran yang berbeda seperti digambarkan di atas. Kelompok Model Analisis Nilai Berbantuan Internet yang mendapatkan enam langkah-langkah pembelajaran berdasarkan sintaknya. Intervensi Model Analisis Nilai Berbantuan Internet meliputi enam langkah pembelajaran Model Analisis Nilai dengan tiga pengujian keputusan. Sedangkan kelompok pembelajaran berbasis perkuliahan tradisional mahasiswa melakukan kegiatan pembelajaran sebagaimana biasa mereka lakukan.

3.7. Sumber Data dan Teknik Pengumpulan Data

Berikut ini dikemukakan Sumber data yang digunakan dalam penelitian ini beserta teknik pengumpulan datanya.

3.7.1 Sumber Data

Sumber data dari mahasiswa pengikut matakuliah Pendidikan Pancasila tahun ajaran 2015/2016 di Universitas Negeri Malang. Data kemampuan berpikir kritis diperoleh dengan data keterampilan berpikir kritis yang menggunakan *California Critical Thinking Skills Test (2000)* dan disposisi berpikir kritis yang menggunakan *California Critical Thinking Disposition Inventory*. Hasilnya tes diolah oleh *Insight Assessment* dan hasilnya digunakan oleh peneliti untuk kepentingan penelitian ini.

3.7.2.Teknik Pengumpulan Data

Data dikumpulkan dari mahasiswa pengikut mata kuliah Pendidikan Pancasila dengan tes tulis yang didatangkan langsung dari California tempat instrumen ini dikeluarkan.. Semua mahasiswa diberi informasi tentang adanya tes CCTST dan CCTDI. Kemudian, peneliti mengumpulkan data dari seluruh peserta. Seluruh peserta menempuh tes sesuai waktu yang ditentukan. Batasan waktunya yang ditetapkan adalah 45 menit untuk CCTST dan 30 menit untuk CCTDI.

3.8. Metode Analisis Data

Pengolahan hasil tes CCTST dan CCTDI dilakukan oleh *Insight Assessment*. Pengkategorian untuk menafsirkan hasil skor didasarkan pada Rekomendasi Asesmen Kinerja yang dibuat oleh penerbit. Data tersebut dianalisis dengan menggunakan SPSS 21. Ciri-ciri umum ditentukan oleh mean statistic deskriptif dan untuk menjawab hubungan variabel dari pertanyaan penelitian diuji kebermaknaan perbedaan skor digunakan uji t sampel independen dan *Analysis of Varian (ANOVA)*.

Independent Sampel T-Tests jika diterjemahkan dalam bahasa Indonesia berarti uji t sampel tidak berhubungan atau bebas. Artinya tidak ada hubungan antara dua sampel yang akan diuji. Dalam penelitian ini akan menguji apakah ada perbedaan rata-rata dua sampel yang terdiri dari kelompok eksperimen (Model Analisis Nilai Berbantuan Internet) dengan kelompok kontrol (menggunakan

model konvensional) terkait dengan keterampilan berpikir kritis maupun pengujian terhadap disposisi berpikir kritis. Jadi kedua sampel penelitian ini terdiri dari dua kelompok yang berbeda atau tidak berhubungan satu sama lain (sampel independen).

Selanjutnya ketika akan menguji variabel jenis kelamin, asal jurusan, ataupun nasal daerah yang berarti ada empat kelompok (yaitu kombinasi antara jenis kelamin dan model pembelajaran, asal jurusan dan model pembelajaran, ataupun asal daerah dan model pembelajaran) digunakanlah Analisis varians/ragam *(analysis of variance*, ANOVA). Analisis varians/ragam *(analysis of variance, ANOVA)* adalah suatu metode analisis statistika yang termasuk ke dalam cabang statistika inferensi yang akan menguji tiga kelompok atau lebih empat kelompok. Dalam penelitian ini ada empat kelompok.

Dengan membandingkan varians/ragam. itu kita bisa mengetahui apakah terdapat perbedaan atau tidak. Perbandingan antar varians ini merupakan uji f. Dalam *analysis of variance* hanya satu hipotesis yang digunakan yaitu hipotesis dua arah *(two tail).* artinya hipotesisnya apakah ada perbedaan rata-rata, yang tidak spesifik kelompok yang mana yang berbeda. Untuk mengetahui kelompok yang benar-benar terdapat perbedaan rata-rata dilakukan uji lanjutan. Berikut hipotesis dalam ANOVA. H0: $\mu 1 = \mu 2 = \mu 3 = \mu 2$, Tidak ada perbedaan yang nyata antara rata-rata hitung dari empat kelompok

CCTST dan CCTDI dirancang untuk memberikan ukuran objektif mengenai kekuatan dan kelemahan dari aspek penting berpikir. Laporan dari penerbit CCTST dan CCTDI memberikan *hasil individu dan kelompok* yang mencakup berbagai informasi statistik dan demografis tentang individu dan kelompok peserta tes. Ringkasan skor peserta tes dan kelompok dan ringkasan kelompok disajikan dengan analisis interpretatif oleh pakar pengukuran *Insight Assessment*.

3.8.1. Analisis Skor Individu dan Kelompok

Berikut ini dikemukakan cara menganalisis skor yang diperoleh berdasarkan skor individu dan kelompok.

Analisis Peserta Tes Individu:

1) skor keseluruhan kemampuan berpikir (Skor Total)

2) interpretasi kategoris kekuatan dari skor keseluruhan dan skor-skor skala
3) peringkat persentil berbasis norma *(a norm-referenced percentile ranking)*
4) skor subskala mengindikasikan bidang keterampilan mana yang secara khusus kuat dan lemah yang perlu perhatian untuk perbaikan.

Untuk menguji kelompok:

1) statistik deskriptif dan representasi grafis dari mean skor total dan skor subskala untuk kelompok
2) grafis dan representasi deskriptif mencakup ukuran kelompok, mean, median, standar deviasi, standar error dari mean, skor terendah, skor tertinggi, skor kuartil pertama dan skor kuartil ketiga.
3) statistik deskriptif dari karakteristik demografi kelompok peserta tes.
4) persentil rata-rata kelompok *(the average percentile of group)* dibandingkan dengan kelompok norma eksternal yang dipilih sebelumnya (dalam hal ini mahasiswa S1 empat tahun).

Skor hasil tes yang diperoleh dari *Insight Assessment* ini kemudian dijadikan bahan analisis untuk menjawab pertanyaan penelitian dalam disertasi ini. Analisis data untuk menjawab pertanyaan penelitian dilakukan dengan uji t sampel independent untuk menguji dua kelompok (misalnya, kelompok kontrol dan eksperimen), dan uji Anova untuk sampel lebih dari dua (misalnya, variabel gender yang membagi peserta tes menjadi empat: laki-laki kelompok kontrol, perempuan kelompok kontrol, perempuan kelompok eksperimen dan laki-laki kelompok eksperimen, demikian juga dengan variabel asal jurusan, asal daerah, tingkat sosial ekonomi). Untuk memperjelas tampilan data digunakan statistik deskriptif dalam bentuk tabel dan tampilan gambar grafis.

Untuk pengolahan dan analisis data digunakan bantuan IBM SPSS 21 dan Excell 2016. Rumus yang digunakan tidak ditampilkan dalam tulisan ini, karena rumus itu sudah lazim digunakan dan sudah termasuk dalam *software* yang digunakan.

Analisis data dilakukan pada masing-masing kelompok (baik dalam pre maupun pos tes), performa mahasiswa pada CCTST dan CCTDI dihitung dan

dikumpulkan berdasarkan kelompok. Perbedaan dalam rata-rata antara kelompok perlakuan dan kelompok pembanding dianalisis dengan uji t, anova sesuai dengan pertanyaan penelitian. Penggunan Data paket IBM SPSS versi 21 dan excel secara terpadu untuk analisis guna *cross-check* hasil analisis dan digunakan untuk melengkapi penyajian data. Statistik deskriptif digunakan untuk menganalisa data demografi peserta. Untuk skor berpikir kritis, mean, standar penyimpangan (SD) dan seluruh analisis item dihitung untuk masing-masing kelompok. Selain itu, Model Analisis Nilai Berbantuan Internet dibandingkan dengan yang konvensional pada CCTST dihitung dengan menggunakan uji t dan anova. Tingkat signifikan ditetapkan pada 0,05 sebagaimana lazimnya penelitian dalam dunia pendidikan.

3.8.2. Menginterpretasikan Laporan Skor Peserta Tes

Setiap tes yang diambil menyediakan empat jenis informasi tentang peserta tes: Skor keseluruhan kemampuan berpikir kritis, rekomendasi asesmen kinerja kekuatan skor total, peringkat persentil dari skor ini bila dibandingkan dengan kelompok peserta tes yang sama, dan satu set skor skala yang membantu untuk memahami bidang keterampilan yang mana yang secara khusus sangat kuat dan yang lebih lemah dan membutuhkan perhatian.

Langkah-langkah berikut digunakan untuk memeriksa setiap jenis informasi karena masing-masing dapat memberikan wawasan tentang kekuatan keterampilan berpikir kritis mahasiswa. Langkah ini juga informatif untuk menginterpretasikan skor mahasiswa individu maupun skor kelompok.

Langkah-langkah Analisis data skor pada CCTST dan CCTDI:

Langkah 1: Menginterpretasikan Skor Total Individu

Menginterpretasikan nilai dari skor total dan informasi yang diterima dari penerbit tentang pengujian perilaku. Skor total CCTST adalah alat ukur keseluruhan keterampilan berpikir kritis terbaik untuk membandingkan masing-masing mahasiswa pada instrumen.

Langkah 2: Mengkaji Skor Norma Persentil Individual

Mengkaji ranking persentil yang merepresentasikan mean skor persentil dari masing-masing anggota kelompok yang diuji. Skor persentil membandingkan peserta tes dengan patokan eksternal kelompok pembanding norma yang paling

mirip dengan sampel ujinya (skor penelitian ini dibandingkan dengan kelompok mahasiswa S1). Dalam sampel peserta tes ada pada jajaran skor keseluruhan CCTST dan berbagai peringkat persentil yang sesuai. Jika menggunakan skor bagian persentil lembaga kita untuk CCTST itu, referensinya skor norma persentil yang dilaporkan. (Dalam CCTDI tidak ada skor persentil norma individual).

Contoh 1 Interpretasi *Percentile Norm*: persentil norma, dalam kolom ditandai Persentil. Sebuah skor yang berada di persentil ke-60 berarti bahwa sekitar 59 orang dari 100 akan mendapat skor lebih rendah dari peserta tes ini dan 40 orang dari 100 akan mendapat skor lebih tinggi daripada peserta tes ini dalam kelompok pembanding keseluruhan.

Langkah 3: Menentukan Kekuatan Skor Total CCTST dengan menggunakan Tabel Rekomendasi Asesmen Kinerja

Dengan memeriksa skor untuk Skor Total Keterampilan Berpikir CCTST yang dibandingkan dengan rekomendasi asesmen kinerja dapat ditentukan kekuatan setiap mahasiswa dan kelompoknya.

Skor bagian Numerik yang Direkomendasikan untuk Rekomendasi Asesmen Kinerja dari Skor Keseluruhan CCTST

Pada Tabel 3.2 disajikan skor numerik CCTST (2000, skala 34-point) versi tingkat perguruan tinggi untuk menentukan rekomendasi asesmen kinerja dan pada Tebel 3.5 menampilkan analisis interpretatif skor CCTDI.

Langkah 4: Menginterpretasi Skor Subskala CCTST

Tujuan dari skor subskala CCTST adalah untuk mengidentifikasi bidang kekuatan pada individu dan bidang kelemahan relatif yang harus ditangani dalam kesempatan pendidikan berikutnya.

Untuk menginterpretasi Skor CCTST, hasil tes Skor Total CCTST dikonsultasikan dengan Tabel 3.2 Rekomendasi Asesmen Kinerja Skor Total CCTST dan untuk CCTST *Skor Skala* dengan Tabel 3.3 Rekomendasi Asesmen Kinerja *Skor Skala* CCTST.

Untuk menginterpretasi Skor CCTDI, hasil tes Skor Total CCTDI dikonsultasikan dengan Tabel 3.4 Rekomendasi Asesmen Kinerja Skor Skor Skala CCTDI

Langkah 5: Mengolah data dari penerbit instrumen untuk menjawab pertanyaan penelitian dengan uji t dan Anova yang dibantu dengan software IBM SPSS 21 dan Excell 1016.

Pertimbangan Penting Ketika Menganalisis Laporan Skor

Peningkatan Individu: Ketika mahasiswa yang sama telah menempuh tes pada dua titik waktu (sebelum dan setelah treatmen) akan dapat diukur peningkatannya dengan menguji perbedaan skor (Posttest – Pretest).

Peningkatan dalam Hubungan dengan Ukuran Sampel: Ukuran sampel merupakan faktor penting dalam analisis statistik. Peningkatan yang lebih besar diperlukan untuk signifikansi statistik dalam sampel kecil. Peningkatan kelompok dua poin signifikan secara edukatif untuk keseluruhan kelompok, kemungkinan merupakan peningkatan yang sangat signifikan pada beberapa individu dalam kelompok. Jika kurang dari 30 orang dalam kelompok, uji statistik akan melaporkan rentangan peningkatan ini sebagai tidak signifikan secara numerik. Penelitian ini menggunakan sampel yang besar 184 orang.

Kerepresentatifan: Perlu hati-hati menggeneralisasikan hasil sampel kecil terhadap asumsi tentang populasi secara keseluruhan, kecuali sampel peserta tes merupakan perwakilan dari populasi yang lebih besar. Sebagai contoh, hasil tes dari sampel 100 mahasiswa belum tentu mewakili populasi yang lebih besar dari mahasiswa. Demikian pula, skor tes dari sampel mahasiswa mahasiswa perkotaan mungkin tidak mewakili populasi yang lebih besar dari mahasiswa jika kelompok yang lebih besar ini termasuk peserta didik dari desa.

Agregat *vs Matched-pairs Comparisons:* Perbandingan kelompok dapat dianalisis secara statistik dalam pendekatan perbandingan cocok-berpasangan (yang mengaitkan skor posttest masing-masing individu dengan skor pretestnya) atau ketika kelompok tidak terdiri dari seperangkat individu yang sama persis, sebagai agregasi skor. Bila mungkin, gunakan pendekatan cocok-berpasangan untuk perbandingan pretest posttest.

Membuang Tes yang Salah: Skor rata-rata kelompok CCTST digunakan untuk mendokumentasikan skor rata-rata untuk kelompok setelah pelaksanaan strategi pembelajaran. Beberapa individu yang tidak memiliki motivasi internal yang memadai untuk terlibat tes yang menantang secara kognitif dengan upaya

tulus ketika mereka tidak melihat keuntungan pribadi (ini terutama bila tes anonym). Alih-alih menolak untuk diuji, kondisi program pendidikan dapat mengarahkan pada perolehan tes palsu (penyelesaian ujian yang tergesa-gesa, sejumlah besar pertanyaan yang tidak terjawab).

1) Skor total yang sangat rendah dianggap sebagai skor yang benar kecuali skor itu dinilai sebagai skor palsu. Skor sangat rendah menunjukkan bahwa peserta tes memiliki keterampilan berpikir kritis sangat lemah, dan ini tidak boleh dibuang karena merepresentasikan skor benar dalam sampel Anda. Namun, skor ujian yang berada di kisaran persentil terendah jika dibandingkan dengan kelompok norma (persentil 0-5) dapat dikaji sebagai kemungkinan tes palsu. Beberapa kemungkinan alasan mengapa skor total sangat rendah mungkin palsu mencakup masalah kemahiran bahasa atau gangguan di pusat pengujian.

2) Hasil dari waktu tes kurang dari 20 menit atau rasio item diselesaikan kurang dari 0,60 harus diperiksa sebagai kemungkinan tes palsu. Dalam sistem pengujian elektronik *Insight Assessment*, tes yang dijawab kurang dari 60% dari pertanyaan yang dijawab dibuang dari kelompok.

3) Reliabilitas uji tes ulang sangat tinggi untuk CCTST. Jika tidak ada faktor intervening, seperti strategi pembelajaran yang memang bertujuan untuk meningkatkan keterampilan berpikir kritis, peserta tes kemungkinan mendapat skor yang sama atau sebanding dari pretestnya ketika diuji ulang.

4) Skor yang turun drastis pada posttest perlu penjelasan. Keterampilan berpikir kritis tidak memburuk selama periode waktu yang singkat, kecuali ada cedera kognitif yang mengintervensi, sehingga pengamatan terhadap penurunan yang signifikan dalam skor total dari pretest ke posttest (skor posttest - skor pretest = skor perbedaan) pada individu tertentu merupakan indikator tes palsu. Perbedaan skor yang wajar adalah antara skor sama hingga atau kurang dari - 3).