

INSTRUMEN PENELITIAN

A. Teknik Pengumpulan Data

Kegiatan penelitian yang terpenting adalah pengumpulan data. Menyusun instrumen adalah pekerjaan penting di dalam langkah penelitian, tetapi mengumpulkan data jauh lebih penting lagi, terutama jika peneliti menggunakan metode yang rawan terhadap masuknya unsur subjektif peneliti. Itulah sebabnya menyusun instrumen pengumpulan data harus ditangani secara serius agar diperoleh hasil yang sesuai dengan kegunaannya yaitu pengumpulan variabel yang tepat.

Pengumpulan data dalam penelitian perlu dipantau agar data yang diperoleh dapat terjaga tingkat validitas dan reliabilitasnya. Walaupun telah menggunakan instrumen yang valid dan reliabel tetapi jika dalam proses penelitian tidak diperhatikan bisa jadi data yang terkumpul hanya ongkongan sampah. Peneliti yang memiliki jawaban responden sesuai keinginannya akan semakin tidak reliabel. Petugas pengumpulan data yang mudah dipengaruhi oleh keinginan pribadinya, akan semakin condong (bias) data yang terkumpul. Oleh karena itu, pengumpul data walaupun tampaknya hanya sekedar pengumpul data tetapi harus tetap memenuhi persyaratan tertentu yaitu yang mempunyai keahlian yang cukup untuk melakukannya.

Mengumpulkan data memang pekerjaan yang melelahkan dan sulit. Dalam penelitian sosial, bisa jadi petugas pengumpul data berjalan dari sekolah ke sekolah dan atau dari rumah ke rumah mengadakan interviu atau membagi angket. Suatu saat terkadang sangat mudah menemukan responden tetapi pada saat yang lain sangat sulit sehingga menimbulkan keputus asaan. Karena itu terkadang pekerjaan pengumpul data seperti sering diberikan kepada pembantu-pembantu peneliti yunior, sedangkan para senior cukup membuat desain, menyusun instrumen, mengolah data, dan mengambil kesimpulan. Seperti sudah dijelaskan, data yang diungkap dalam penelitian dapat dibedakan menjadi tiga jenis, yaitu: fakta, pendapat, dan kemampuan. Untuk mengukur ada atau tidaknya atau besar

kecilnya kemampuan objek yang diteliti, seringkali menggunakan tes. Perlu kita ketahui, pelaksanaan tes bukan hanya untuk mengukur kemampuan manusia tetapi tes dapat juga dilakukan untuk mengukur kemampuan mesin atau perlengkapan lainnya. Juga Bahkan seekor binatang seperti anjing pelajar perlu juga di-tes. Dari test akan diketahui ada yang memiliki kemampuan yang rendah dan ada pula yang tinggi.

Untuk manusia, instrumen yang berupa tes ini dapat digunakan untuk mengukur kemampuan dasar dan pencapaian atau prestasi. Untuk mengukur kemampuan dasar antara lain dengan tes inteligensi (IQ), tes minat, tes bakat khusus, dan sebagainya. Khusus untuk tes prestasi belajar yang biasa digunakan di sekolah adalah tes buatan guru dan tes terstandar yang dibuat oleh tim khusus secara nasional dan internasional

1. Pengumpulan data melalui Kuesioner atau Angket Sebagian besar penelitian umumnya menggunakan kuesioner sebagai metode yang dipilih untuk mengumpulkan data. Kuesioner atau angket memang mempunyai banyak kebaikan sebagai instrumen pengumpul data. Prosedur penyusunan kuesioner:
 - a. Merumuskan tujuan yang akan dicapai dengan kuesioner.
 - b. Mengidentifikasi variabel yang akan dijadikan sasaran kuesioner.
 - c. Menjabarkan setiap variabel menjadi sub-variabel yang lebih spesifik dan tunggal.
 - d. Menentukan jenis data yang akan dikumpulkan, sekaligus untuk menentukan teknik analisisnya.

Penentuan sampel sebagai responden kuesioner perlu mendapat perhatian pula. Apabila salah menentukan sampel, informasi yang kita butuhkan barangkali tidak kita peroleh secara maksimal.

2. Pengumpulan data melalui Metode Interview

Penggunaan metode interview memerlukan waktu yang cukup lama untuk mengumpulkan data. Dibandingkan dengan mengedarkan angket kepada responden, interview sangat rumit. Dalam melakukan interview, peneliti harus memperhatikan sikap pada waktu datang, sikap duduk, kecerahan wajah, tutur kata, keramahan, kesabaran serta keseluruhan penampilan, akan sangat berpengaruh terhadap

isi jawaban responden yang diterima oleh peneliti. Oleh sebab itu, maka perlu adanya latihan yang intensif bagi calon *interviewer*.

Secara garis besar ada dua macam pedoman wawancara yaitu pedoman wawancara tidak terstruktur, yaitu pedoman wawancara yang hanya memuat garis besar yang akan ditanyakan. Tentu saja kreativitas pewawancara sangat diperlukan, bahkan hasil wawancara dengan jenis pedoman ini lebih banyak tergantung dan pewawancara. Pewawancara sebagai pengemudi jawaban responden. Jenis wawancara ini cocok untuk penelitian kasus. Dan jenis kedua adalah pedoman wawancara terstruktur, yaitu pedoman wawancara yang disusun secara terperinci sehingga menyerupai check-list. Pewawancara tinggal membubuhkan tanda v (*check*) pada nomor yang sesuai.

Pedoman wawancara yang banyak digunakan adalah bentuk “semi structured”. Dalam hal ini maka mula-mula interviwer menanyakan serentetan pertanyaan yang sudah terstruktur, kemudian satu per satu diperdalam dalam mengorek keterangan lebih lanjut. Dengan demikian jawaban yang diperoleh bisa meliputi semua variabel, dengan keterangan yang lengkap dan mendalam.

3. Pengumpulan data melalui Metode observasi

Dalam menggunakan metode observasi cara yang paling efektif adalah melengkapinya dengan format atau blangko pengamatan sebagai instrumen. Format yang disusun berisi item-item tentang kejadian atau tingkah laku yang digambarkan akan terjadi. Peranan yang paling penting dalam menggunakan metode observasi adalah pengamat. Pengamat harus jeli dalam mengamati adalah menatap kejadian, gerak atau proses. Mengamati bukanlah pekerjaan yang mudah karena manusia banyak dipengaruhi oleh minat dan kecenderungan-kecenderungan yang ada padanya. Padahal hasil pengamatan harus sama, walaupun dilakukan oleh beberapa orang. Dengan lain perkataan, pengamatan harus objektif.

4. Pengumpulan Data melalui Metode Dokumentasi

Tidak kalah penting dan metode-metode lain, adalah metode dokumentasi, yaitu mencari data mengenai hal-hal atau variabel

yang berupa catatan, transkrip, buku, surat kabar, majalah, prasasti, notulen rapat, lengger, agenda, dan sebagainya. Dibandingkan dengan metode lain, maka metode ini agak tidak begitu sulit, dalam arti apabila ada kekeliruan sumber datanya masih tetap, belum berubah. Dengan metode dokumentasi yang diamati bukan benda hidup tetapi benda mati. Seperti telah dijelaskan, dalam menggunakan metode dokumentasi ini peneliti memegang cek-list untuk mencari variabel yang sudah ditentukan. Apabila terdapat/muncul variabel yang dicari, maka peneliti tinggal membubuhkan tanda check atau tally di tempat yang sesuai. Untuk mencatat hal-hal yang bersifat bebas atau belum ditentukan dalam daftar variabel peneliti dapat menggunakan kalimat bebas.

B. Instrumen Penelitian

Menyusun instrumen merupakan langkah penting dalam pola prosedur penelitian. Instrumen berfungsi sebagai alat bantu dalam mengumpulkan data yang diperlukan. Bentuk instrumen berkaitan dengan metode pengumpulan data, misal metode wawancara yang instrumennya pedoman wawancara. Metode angket atau kuesioner, instrumennya berupa angket atau kuesioner. Metode tes, instrumennya adalah soal tes, tetapi metode observasi, instrumennya bernama cek-list (Black, 2006).

Menyusun instrumen pada dasarnya adalah menyusun alat evaluasi, karena mengevaluasi adalah memperoleh data tentang sesuatu yang diteliti, dan hasil yang diperoleh dapat diukur dengan menggunakan standar yang telah ditentukan sebelumnya oleh peneliti. Dalam hal ini terdapat dua macam alat evaluasi yang dapat dikembangkan menjadi instrumen penelitian, yaitu tes dan non-tes (Narbuko, 2004).

1. Bentuk Instrumen Tes

Tes dapat berupa serentetan pertanyaan, lembar kerja, atau sejenisnya yang dapat digunakan untuk mengukur pengetahuan, keterampilan, bakat, dan kemampuan dari subjek penelitian. Lembar instrumen berupa tes ini berisi soal-soal tes yang terdiri atas butir-butir soal. Setiap butir soal mewakili satu jenis variabel yang diukur.

Berdasarkan sasaran dan objek yang diteliti, terdapat beberapa macam tes, yaitu:

- a. Tes kepribadian atau personality test, digunakan untuk mengungkap kepribadian seseorang yang menyangkut konsep pribadi, kreativitas, disiplin, kemampuan, bakat khusus, dan sebagainya,
- b. Tes bakat atau aptitude test, tes ini digunakan untuk mengetahui bakat seseorang,
- c. Tes inteligensi atau intelligence test, dilakukan untuk memperkirakan tingkat intelektual seseorang,
- d. Tes sikap atau attitude test, digunakan untuk mengukur berbagai sikap orang dalam menghadapi suatu kondisi,
- e. Tes minat atau measures of interest, ditujukan untuk menggali minat seseorang terhadap sesuatu,
- f. Tes prestasi atau achievement test, digunakan untuk mengetahui pencapaian seseorang setelah dia mempelajari sesuatu.

Bentuk instrumen ini dapat dipergunakan salah satunya dalam mengevaluasi kemampuan hasil belajar siswa di sekolah dasar, tentu dengan memperhatikan aspek aspek mendasar seperti kemampuan dalam pengetahuan, sikap serta keterampilan yang dimiliki baik setelah menyelesaikan salah satu materi tertentu atau seluruh materi yang telah disampaikan.

2. Bentuk Instrumen Angket atau Kuesioner

Angket atau Kuesioner adalah metode pengumpulan data, instrumennya disebut sesuai dengan nama metodenya. Bentuk lembaran angket dapat berupa sejumlah pertanyaan tertulis, tujuannya untuk memperoleh informasi dari responden tentang apa yang ia alami dan ketahuinya.

Bentuk kuesioner yang dibuat sebagai instrumen sangat beragam, seperti:

- a. Kuesioner terbuka, responden bebas menjawab dengan kalimatnya sendiri, bentuknya sama dengan kuesioner isian.
- b. Kuesioner tertutup, responden tinggal memilih jawaban yang telah disediakan, bentuknya sama dengan kuesioner pilihan ganda
- c. Kuesioner langsung, responden menjawab pertanyaan seputar

dirinya

- d. Kuesioner tidak langsung, responden menjawab pertanyaan yang berhubungan dengan orang lain
- e. Check list, yaitu daftar isian yang bersifat tertutup, responden tinggal membubuhkan tanda check pada kolom jawaban yang tersedia
- f. Skala bertingkat, jawaban responden dilengkapi dengan pernyataan bertingkat, biasanya menunjukkan skala sikap yang mencakup rentang dari sangat setuju sampai sangat tidak setuju terhadap pernyataannya.

Setelah bentuk kuesioner ditetapkan, langkah selanjutnya adalah membuat pertanyaan dengan mempertimbangkan jumlah pertanyaan agar tidak terlalu banyak atau terlalu sedikit, yang penting disesuaikan dengan indikator yang ditetapkan. Kemudian tidak menanyakan hal yang tidak perlu semisal nomor telp responden yang jelas tidak akan di oleh dalam penelitian.

Dalam menata tampilan pada lembar kuesioner, perlu diperhatikan hal-hal yang berkaitan dengan keindahan, kemudahan mengisi, dan kemudahan memeriksa jawaban. Oleh karena itu diperlukan kreativitas untuk membuat tampilan kuesioner menjadi enak dibaca, seperti penggunaan garis-garis dan kotak pada hal-hal yang dianggap penting, penggunaan warna-warna dan hiasan, serta meletakkan kelompok pertanyaan tentang identitas pengisi, pengantar, dan pertanyaan inti pada tempat yang berbeda.

Bentuk tes seperti ini dapat saudara laksanakan salah satunya ketika menyelesaikan tugas akhir terkait dengan bidang garapan ke SD an diantaranya membuat laporan tugas akhir penyelesaian studi seperti skripsi.

3. Bentuk Instrumen Interview

Suatu bentuk dialog yang dilakukan oleh peneliti untuk memperoleh informasi dari responden dinamakan interview. Instrumennya dinamakan pedoman wawancara atau inter view guide. Dalam pelaksanaannya, interview dapat dilakukan secara bebas artinya pewawancara bebas menanyakan apa saja kepada

terwawancara tanpa harus membawa lembar pedomannya. Syarat interviu seperti ini adalah pewawancara harus tetap mengingat data yang harus terkumpul.

Lain halnya dengan interviu yang bersifat terpimpin, pewawancara berpedoman pada pertanyaan lengkap dan terperinci, layaknya sebuah kuesioner. Selain itu ada juga interviu yang bebas terpimpin, dimana pewawancara bebas melakukan interviu dengan hanya menggunakan pedoman yang memuat garis besarnya saja.

Kekuatan interviu terletak pada keterampilan seorang interviewer dalam melakukan tugasnya, dia harus membuat suasana yang tenang, nyaman, dan bersahabat agar sumber data dapat memberikan informasi yang jujur. Si interviewer harus dibuat terpancing untuk mengeluarkan informasi yang akurat tanpa merasa diminta secara paksa, ibaratnya informasi keluar seperti air mengalir dengan deras.

Tes ini sangat tepat dilakukan oleh peneliti yang ingin mendapatkan informasi terkini terkait dengan berbagai kejadian, seperti ketika seorang guru sekolah dasar ingin mendapatkan gambaran menyeluruh tentang kinerja salah seorang guru di sekolah tertentu, maka lakukan dengan wawancara diantaranya dengan kepala sekolah, dengan teman sejawat serta wawancara dilakukan dengan sebagian siswa yang telah mengikuti kegiatan pembelajaran dengan guru terkait.

4. Bentuk Instrumen Observasi

Observasi dalam sebuah penelitian diartikan sebagai pemusatan perhatian terhadap suatu objek dengan melibatkan seluruh indera untuk mendapatkan data. Jadi observasi merupakan pengamatan langsung dengan menggunakan penglihatan, penciuman, pendengaran, perabaan, atau kalau perlu dengan pengecapan. Instrumen yang digunakan dalam observasi dapat berupa pedoman pengamatan, tes, kuesioner, rekaman gambar, dan rekaman suara.

Instrumen observasi yang berupa pedoman pengamatan, biasa digunakan dalam observasi sistematis dimana si pelaku observasi bekerja sesuai dengan pedoman yang telah dibuat. Pedoman

tersebut berisi daftar jenis kegiatan yang kemungkinan terjadi atau kegiatan yang akan diamati. Sebagai contoh, observasi yang dilakukan di sebuah sekolah, objek yang akan diamati ditulis dalam pedoman tersebut secara berurutan dalam sebuah kolom yang akan di tally, isi daftarnya adalah berbagai peristiwa yang mungkin terjadi di sekolah tersebut seperti: kepala sekolah memberi pengarahan kepada guru-guru, guru piket mengisi materi pada kelas yang pengajarnya berhalangan hadir, petugas administrasi mengisi buku induk siswa, penjaga sekolah memelihara peralatan kebersihan sekolah, murid-murid berseragam rapih, dan sebagainya. Bekerja dengan pedoman pengamatan seperti ini dinamakan sistem tanda (*sign system*), data yang didapatkan berupa gambaran singkat (*snapshot*) mengenai situasi warga sekolah dalam suatu hari tertentu (Roberts, dkk, 2002; Popay, 2006).

Ada lagi satu bentuk instrumen observasi yang dinamakan *category system*, yaitu sistem pengamatan yang membatasi pada sejumlah variabel. Hal yang diamati terbatas pada kejadian-kejadian yang termasuk dalam kategori variabel, di luar itu, setiap kejadian yang berlangsung tidak diamati atau diabaikan saja. Contoh, pengamatan terhadap kinerja kepala sekolah, maka kejadian yang diamati dan ditally adalah kepala sekolah datang ke sekolah tepat waktu, kepala sekolah mengamati proses belajar mengajar, kepala sekolah membuat rancangan program peningkatan kualitas guru dan murid, dan sebagainya. Hasil pengamatan menyimpulkan bahwa kepala sekolah tersebut memiliki kinerja yang baik atau buruk.

Selain bentuk instrumen berupa pedoman pengamatan, terdapat juga instrumen observasi dalam bentuk tes yang digunakan untuk mengamati aspek kejiwaan. Kemudian bentuk kuesioner yang diberikan kepada responden untuk mengamati aspek-aspek yang ingin diselidiki, dan rekaman gambar serta rekaman suara yang digunakan sebagai penyimpan sumber data, dimana sumber data dapat diamati lebih lama bahkan berulang-ulang sesuai kebutuhan (Sevilla, dkk, 1993).

5. Bentuk Instrumen Skala Bertingkat atau Rating Scale

Bentuk instrumen dengan skala bertingkat lebih memudahkan peneliti untuk mengetahui pendapat responden lebih mendalam tentang variabel yang diteliti. Rating atau skala bertingkat adalah suatu ukuran subjektif yang dibuat berskala. Yang harus diperhatikan dalam pembuatan rating scale adalah kehati-hatian dalam membuat skala, agar pernyataan yang diskalakan mudah diinterpretasi dan responden dapat memberikan jawaban secara jujur.

Untuk mengantisipasi ketidakjujuran jawaban dari responden, maka perlu diwaspadai beberapa hal yang mempengaruhinya. Menurut Gall, dkk (2003) faktor yang berpengaruh terhadap ketidakjujuran jawaban responden adalah a) persahabatan, (b) kecepatan menerka, (c) cepat memutuskan, (d) jawaban kesan pertama, (e) penampilan instrumen, (f) prasangka, (g) halo effects, (h) kesalahan pengambilan rata-rata, dan (i) kemurahan hati.

6. Bentuk Instrumen Dokumentasi

Bentuk instrumen dokumentasi terdiri atas dua macam yaitu pedoman dokumentasi yang memuat garis-garis besar atau kategori yang akan dicari datanya, dan check-list yang memuat daftar variabel yang akan dikumpulkan datanya. Perbedaan antara kedua bentuk instrumen ini terletak pada intensitas gejala yang diteliti. Pada pedoman dokumentasi, peneliti cukup menuliskan tanda centang dalam kolom gejala, sedangkan pada check-list, peneliti memberikan tally pada setiap pemunculan gejala (Cooper, dkk, 2002).

Instrumen dokumentasi dikembangkan untuk penelitian dengan menggunakan pendekatan analisis isi. Selain itu digunakan juga dalam penelitian untuk mencari bukti-bukti sejarah, landasan hukum, dan peraturan-peraturan yang pernah berlaku. Subjek penelitiannya dapat berupa buku-buku, majalah, dokumen, peraturan-peraturan, notulen rapat, catatan harian, bahkan benda-benda bersejarah seperti prasasti dan artefak (Clemmens, 2003).

C. Validitas Instrumen

Persoalan alat ukur yang digunakan evaluator ketika melakukan kegiatan evaluasi sering dihadapkan pada persoalan akurasi, konsisten dan stabilitas sehingga hasil pengukuran yang diperoleh bisa mengukur dengan

akurat sesuatu yang sedang diukur. Instrumen ini memang harus memiliki akurasi ketika digunakan. Konsisten dan stabil dalam arti tidak mengalami perubahan dari waktu pengukuran satu ke pengukuran yang lain (Ali, 1993; Anggoro, 2008).

Alat ukur atau instrumen yang baik harus memenuhi dua syarat yaitu validitas dan reliabilitas. Suatu alat ukur yang tidak reliabel atau tidak valid akan menghasilkan kesimpulan yang bias, kurang sesuai dengan yang seharusnya, dan akan memberikan informasi yang keliru mengenai keadaan subjek atau individu yang dikenai tes itu. Apabila informasi yang keliru itu dengan sadar atau tidak dengan sadar digunakan sebagai dasar pertimbangan dalam pengambilan suatu keputusan, maka keputusan itu tentu bukan merupakan suatu keputusan yang tepat (Bryman, 2004).

Alat ukur atau instrumen yang akan disusun tentu saja harus memiliki validitas dan reliabilitas, agar data yang diperoleh dari alat ukur itu bisa reliabel, valid dan disebut dengan validitas dan reliabilitas alat ukur atau validitas dan reliabilitas instrumen.

Validitas adalah salah satu ciri yang menandai tes hasil belajar yang baik. Untuk dapat menentukan apakah suatu tes hasil belajar telah memiliki validitas atau daya ketepatan mengukur, dapat dilakukan dari dua segi, yaitu : dari segi tes itu sendiri sebagai totalitas, dan dari segi itemnya, sebagai bagian yang tak terpisahkan dari tes tersebut (Bloor, 1997). Di dalam buku "Encyclopedia of Educational Evaluation," Scarvia B. Anderson mengatakan bahwa "A test is valid if it measures what it purpuse to measure" artinnya : "sebuah tes dikatakan valid apabila tes tersebut mengukur apa yang hendak diukur" (Dixon, dkk, 2004).

Penganalisisan terhadap tes hasil belajar sebagai suatu totalitas dapat dilakukan dengan dua cara. Pertama, penganalisisan yang dilakukan dengan jalan berpikir secara rasional atau penganalisisan dengan menggunakan logika (logical analysis). Kedua, penganalisisan yang dilakukan dengan mendasarkan diri kepada kenyataan empiris, dimana penganalisisan dilaksanakan dengan menggunakan empirical analysis (Dowie, 2006).

1. Macam-Macam Validitas

Secara umum, validitas tes dibagi menjadi dua yaitu validitas tes secara rasional dan validitas tes secara empiris.

a. Validitas Tes Secara Rasional

Validitas rasional adalah validitas yang diperoleh atas dasar hasil pemikiran, validitas yang diperoleh dengan berfikir secara logis. Dengan demikian maka suatu tes hasil belajar dapat dikatakan telah memiliki validitas rasional, apabila setelah dilakukan penganalisisan secara rasional ternyata bahwa tes hasil belajar itu memang (secara rasional) dengan tepat telah dapat mengukur apa yang seharusnya diukur (Hermawan, 2005).

Untuk dapat menentukan apakah tes hasil belajar sudah memiliki validitas rasional ataukah belum, dapat dilakukan penelusuran dari dua segi, yaitu :

i. Validitas Isi (*Content Validity*)

Validitas isi artinya kejituan daripada suatu tes ditinjau dari isi tes tersebut. Suatu tes hasil belajar dapat dikatakan valid, apabila materi tes tersebut betul-betul merupakan bahan-bahan yang representatif terhadap bahan-bahan pelajaran yang diberikan. Misalnya apabila kita ingin memberikan tes bahasa Inggris kepada siswa kelas II, maka item-itemnya harus diambil dari bahan-bahan pelajaran kelas II. Apabila terdapat bahan-bahan pelajaran kelas III, maka tes tersebut sudah tidak valid lagi (Sekaran, 2007).

Dalam praktik, validitas isi dari suatu tes hasil belajar dapat diketahui dengan jalan membandingkan antara isi yang terkandung dalam tes hasil belajar, dengan tujuan intruksional khusus yang telah ditentukan untuk masing-masing mata pelajaran, apakah hal-hal yang tercantum dalam tujuan instruksional khusus yang sudah terwakili secara nyata dalam tes hasil belajar tersebut ataukah belum. Jika penganalisisan secara rasional itu menunjukkan hasil yang membenarkan tentang telah tercerminnya tujuan instruksional khusus itu di dalam tes hasil belajar, maka tes hasil belajar yang sedang di uji validitas isinya itu dapat dinyatakan sebagai tes hasil

belajar yang telah memiliki validitas isi (Suprayogo, 2001).

ii. Validitas Konstruk (*Construct Validity*)

Secara etimologis, kata “konstruksi” mengandung arti susunan, kerangka atau rekaan. Validitas susunan artinya kejituan daripada suatu tes ditinjau dari susunan tes tersebut. Misalnya kalau kita ingin memberikan tes kecakapan ilmu pasti, kita harus membuat soal yang ringkas dan jelas yang benar-benar akan mengukur kecakapan ilmu pasti, bukan mengukur kemampuan bahasa karena soal itu ditulis secara berkepanjangan dengan bahasa yang mudah dimengerti (Usman & Purnomo, 2008).

Validitas konstruksi dari suatu tes hasil belajar dapat dilakukan penganalisisannya dengan jalan melakukan pencocokan antara aspek-aspek berfikir yang terkandung dalam tes hasil belajar tersebut, dengan aspek-aspek berfikir yang dikehendaki untuk diungkap oleh tujuan instruksional khusus. Jika secara logis hasil penganalisisan itu menunjukkan bahwa aspek-aspek berfikir yang diungkap melalui butir-butir soal tes hasil belajar itu sudah dengan secara tepat mencerminkan aspek-aspek berfikir yang oleh tujuan instruksional khusus diperintahkan untuk diungkap maka tes hasil belajar tersebut dapat dinyatakan sebagai tes hasil belajar yang valid dari susunannya atau telah memiliki validitas konstruksi (Syah, 2010).

b. Validitas tes secara empiris

Validitas empiris adalah validitas yang bersumber pada pengamatan di lapangan. Tes hasil belajar dapat dikatakan telah memiliki validitas empiris apabila didasarkan hasil analisis yang dilakukan terhadap data hasil pengamatan di lapangan, terbukti bahwa hasil tes belajar itu dengan secara tepat telah dapat mengukur hasil belajar yang seharusnya diungkap atau diukur lewat tes hasil belajar tersebut (Arif, 2005).

Untuk menentukan apakah tes hasil belajar sudah memiliki validitas empiris ataukah belum dapat dilakukan penelusuran dari dua segi yaitu:

i. Validitas ramalan (*predictive validity*)

Validitas ramalan artinya ketepatan (kejituan) daripada suatu alat pengukur ditinjau dari kemampuan tes tersebut untuk meramalkan prestasi yang dicapainya kemudian. Misalnya suatu tes hasil belajar dapat dikatakan mempunyai validitas ramalan yang tinggi, apabila hasil yang dicapai oleh anak dalam tes tersebut betul-betul dapat meramalkan sukses tidaknya anak-anak dalam pelajaran-pelajaran yang akan datang (Bailey, 1998).

Suatu tes hasil belajar dapat dinyatakan sebagai tes yang telah memiliki validitas ramalan atau belum dapat ditempuh dengan cara mencari korelasi antara tes hasil belajar yang sedang diuji validitas ramalannya dengan kriterium yang ada. Jika di antara kedua variabel tersebut terdapat korelasi positif yang signifikan maka tes hasil belajar yang sedang diuji validitas ramalannya itu dapat dinyatakan sebagai tes hasil belajar yang telah memiliki daya ramal yang tepat, artinya apa yang telah diramalkan, betul-betul telah terjadi secara nyata dalam praktek (Attree & Milton, 2006).

ii. Validitas bandingan

Validitas bandingan artinya kejituan daripada suatu tes dilihat dari kolerasinya terhadap kecakapan yang telah dimiliki saat ini secara riil. Perbedaan antara validitas ramalan dengan validitas bandingan ialah dilihat dari segi waktunya. Validitas ramalan melihat hubungannya dengan masa yang akan datang, sedangkan validitas bandingan melihat hubungannya dengan masa sekarang (Hamidi, 2004).

Dalam rangka menguji validitas bandingan, data

yang mencerminkan pengalaman yang diperoleh pada masa lalu itu, kita bandingkan dengan data hasil tes yang diperoleh sekarang ini. Jika hasil tes yang ada sekarang ini mempunyai hubungan searah dengan hasil tes berdasarkan pengalaman yang lalu, maka tes yang memiliki karakteristik seperti itu dapat dikatakan telah memiliki bandingan (Krathwohl, 2006)).

Seperti halnya validitas ramalan, maka untuk mengetahui ada atau tidaknya hubungan yang searah antara tes pertama dengan tes berikutnya, dapat digunakan teknik analisis korelasi antara variabel X (tes pertama) dengan variabel Y (tes berikutnya) adalah positif dan signifikan, maka tes tersebut dapat dinyatakan sebagai tes yang telah memiliki validitas bandingan (Leedy, 2004).

2. Validitas Perangkat Soal

a. Validitas empiris butir soal objektif

Pengertian validitas dipakai untuk butir soal dan soal (perangkat soal), karena dikenal validitas butir soal dan validitas perangkat soal. Perangkat soal terdiri atas sejumlah butir soal, validitas perangkat soal ditentukan oleh validitas butir-butir soalnya. Perangkat soal bersifat valid (sahih) bila butir-butir soalnya valid. Berdasarkan penjelasan sebelumnya, ada dua macam validitas yaitu validitas teoritis (isi dan perilaku) dan validitas empiris.

Validitas empiris butir soal dihitung dengan cara statistik korelasi. Validitas butir soal objektif dihitung dengan rumus korelasi point biserial, validitas butir soal uraian dihitung dengan rumus korelasi product moment. Angka korelasi yang diperoleh dengan cara demikian disebut koefisien validitas atau angka validitas butir soal.

Untuk butir soal objektif validitas butir soal dihitung dengan rumus korelasi point biserial antar masing-masing skor butir soal (X_p) dengan skor total (X_t). Dipakai rumus point biserial

karena data yang dikorelasikan adalah data nominal dengan data interval. Data nominal berasal dari skor butir soal, yaitu 1 untuk jawaban benar dan 0 untuk jawaban salah.

Rumus korelasi *point biserial* :

$$r_{pbi} = \frac{\bar{x}_i - \bar{x}_t}{S_t} \sqrt{\frac{p}{1-p}}$$

keterangan :

$\bar{x}_i$ = mean butir yang menjawab benar

$\bar{x}_t$ = mean skor total

S_t = simpangan baku total

p = proporsi yang menjawab benar

b. Validitas atau kesahihan empiris butir soal uraian

Validitas butir soal uraian dihitung dengan rumus product moment, antara skor butir soal (X_p) dengan skor total (X_t). Dipakai product momen karena data yang dikorelasikan adalah data interval dengan data interval.

i. Rumus product moment dengan simpangan

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum X^2)(\sum Y^2)}}$$

Keterangan:

r_{xy} = koefisien korelasi antara variabel x dan variabel y

$\sum xy$ = jumlah perkalian x dan y

X^2 = kuadrat dari x

Y^2 = kuadrat dari y

ii. Rumus product moment angka kasar

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

Sankaa

keterangan:

r_{xy} = koefisien korelasi

$\sum X$ = jumlah skor butir

$\sum Y$ = jumlah skor total

N = jumlah sampel

3. Faktor-Faktor Yang Mempengaruhi Validitas

Banyak faktor yang dapat mempengaruhi hasil tes evaluasi tidak valid. Beberapa faktor tersebut secara garis besar dapat dibedakan menurut sumbernya, yaitu faktor internal dari tes, faktor eksternal tes, dan faktor yang berasal dari siswa yang bersangkutan.

a. Faktor yang berasal dari dalam tes

- 1) Arahan tes yang disusun dengan makna tidak jelas sehingga dapat mengurangi validitas tes.
- 2) Kata-kata yang digunakan dalam struktur instrumen evaluasi, tidak terlalu sulit.
- 3) Item tes dikonstruksi dengan jelas.
- 4) Tingkat kesulitan item tes tidak tepat dengan materi pembelajaran yang diterima siswa.
- 5) Waktu yang dialokasikan tidak tepat, hal ini termasuk kemungkinan terlalu kurang atau terlalu longgar.
- 6) Jumlah item terlalu sedikit sehingga tidak mewakili sampel.
- 7) Jawaban masing-masing item evaluasi bisa diprediksi siswa.

b. Faktor yang berasal dari administrasi dan skor tes

- 1) Waktu pengerjaan tidak cukup sehingga siswa dalam memberikan jawaban dalam situasi tergesa-gesa.
- 2) Adanya kecurangan dalam tes sehingga tidak membedakan antara siswa yang belajar dengan melakukan kecurangan.
- 3) Pemberian petunjuk dari pengawas yang tidak dapat dilakukan pada semua siswa.
- 4) Teknik pemberian skor yang tidak konsisten.
- 5) Siswa tidak dapat mengikuti arahan yang diberikan dalam tes baku.
- 6) Adanya joki (orang lain bukan siswa) yang masuk dalam menjawab item tes yang diberikan.

c. Faktor yang berasal dari jawaban siswa

Seringkali terjadi bahwa interpretasi terhadap item-item tes evaluasi tidak valid, karena dipengaruhi oleh jawaban siswa dari pada interpretasi item-item pada tes evaluasi (Sukardi, 2009).

D. Reliabilitas Instrumen

Reliabilitas merupakan penerjemahan dari kata reliability yang mempunyai asal kata rely yang artinya percaya dan reliabel yang artinya dapat dipercaya. Keterpercayaan berhubungan dengan ketepatan dan konsistensi. Test hasil belajar dikatakan dapat dipercaya apabila memberikan hasil pengukuran hasil belajar yang relatif tetap secara konsisten. Beberapa ahli memberikan batasan reliabilitas. Menurut Azwar (2012), reliabilitas berhubungan dengan akurasi instrumen dalam mengukur apa yang diukur, kecermatan hasil ukur dan seberapa akurat seandainya dilakukan pengukuran ulang. Azwar juga menyatakan reliabilitas sebagai konsistensi pengamatan yang diperoleh dari pencatatan berulang baik pada satu subjek maupun sejumlah subjek.

Kerlinger memberikan batasan tentang reliabilitas yaitu :

- 1) Reliabilitas dicapai apabila kita mengukur himpunan objek yang sama berulang kali dengan instrumen yang sama atau serupa akan memberikan hasil yang sama atau serupa.
- 2) Reliabilitas dicapai apabila ukuran yang diperoleh dari suatu instrumen pengukur adalah ukuran “ yang sebenarnya” untuk sifat yang diukur.
- 3) Reliabilitas dicapai dengan meminimalkan galat pengukuran yang terdapat pada suatu instrumen pengukur.

Jadi, dari berbagai definisi reliabilitas dapat disimpulkan bahwa reliabilitas berhubungan dengan kemampuan alat ukur untuk melakukan pengukuran secara cermat. Reliabilitas merupakan akurasi dan presisi yang dihasilkan oleh alat ukur dalam melakukan pengukuran.

1. Reliabilitas Atau Keandalan Empiris Soal

Secara empirik, tinggi rendahnya reliabilitas ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas. Soal (perangkat soal) yang valid pasti reliabel, tetapi soal yang reliabel belum tentu valid. Oleh karena itu soal yang valid secara teoritis, juga sudah reliabel (andal) secara teoritis. Dengan demikian soal buatan guru yang sudah disusun melalui kisi-kisi, sudah valid secara teoritis juga sudah reliabel secara teoritis. Reliabilitas empiris soal juga dihitung dengan teknik statistik, yaitu dengan cara korelasi. Angka korelasi

yang diperoleh dengan cara ini disebut koefisien reliabilitas atau angka reliabilitas (r_{11} atau r_{tt}) soal. Soal yang baik adalah soal yang mempunyai koefisien reliabilitas lebih dari sama dengan 0,70.

a. Reliabilitas empiris soal objektif

Reliabilitas empiris soal objektif dibagi menjadi tiga macam, yaitu:

1) Koefisien stabilitas

Koefisien stabilitas (*coefficient of stability*) adalah jenis reliabilitas yang diperoleh dengan cara uji coba ulang (test-retest) yaitu dengan memberikan ujian dengan suatu soal kepada sekelompok individu kemudian mengujikan kembali soal tersebut pada kelompok sama pada waktu yang berbeda. Besarnya reliabilitas soal dihitung dengan mencari product moment antara skor hasil uji pertama dengan skor hasil uji kedua. Soal dikatakan reliabel bila koefisien stabilitas r_{11} atau r_{tt} sama atau lebih besar dari 0,70.

2) Koefisien ekuivalen

Koefisien ekuivalen (*coefficient of equivalence*) adalah jenis reliabilitas yang diperoleh dengan cara menguji cobakan dua soal yang paralel pada kelompok sama dan waktu yang sama (equivalence forms method, parallel form method, atau alternate forms method). Jadi dalam hal ini ada dua soal yang paralel, artinya masing-masing soal disusun tersendiri, jumlah butir soal sama, isi dan bentuk sama, tingkat kesukaran sama, waktu serta petunjuk untuk mengerjakan soal juga sama. Skor hasil uji coba kedua soal dikorelasikan dengan rumus product moment untuk menghitung koefisien ekuivalen. Kedua jenis soal yang paralel bersifat reliabel jika angka koefisien ekuivalen yaitu r_{11} atau r_{tt} besar atau sama dengan 0,70.

3) Koefisien konsistensi internal

Koefisien konsistensi internal (*coefficient of internal consistency*) adalah reliabilitas yang diperoleh dengan cara mengujicobakan suatu soal dan menghitung korelasi hasil uji coba dari kelompok yang sama. Ada tiga cara untuk

memperoleh reliabilitas jenis ini yaitu; cara belah dua (split half method), cara Kuder Richardson 20 atau Kuder Richardson 21, dan cara Cronbach khusus untuk soal uraian.

a) Cara belah dua

Pada cara ini, soal diuji cobakan kepada peserta didik dan hasilnya dibelah menjadi dua, yaitu belahan gasal dan belahan genap. Dalam hal ini jumlah butir soal harus genap. Kedua skor hasil belahan dikorelasikan dengan rumus product moment, hasilnya adalah relasi belahan $r_{1/2\ 1/2}$. Setelah ditemukan korelasi belahan, dihitung angka reliabilitas soal dengan rumus Spearman-Brown. Rumus Spearman-Brown adalah sebagai berikut :

$$r_{11} = \frac{(2)(r_{1/2\ 1/2})}{1 + r_{1/2\ 1/2}}$$

Keterangan :

$r_{1/2\ 1/2}$ = korelasi antara skor-skor setiap belahan tes

r = koefisien reliabilitas yang sudah disesuaikan

Selain dengan rumus Spearman-Brown, dapat pula dipakai rumus Flanagan. Diperlukan data simpangan baku skor belahan gasal (SB_{gasal}), simpangan baku skor belahan genap SB_{genap} dan simpangan baku skor total SB_{total} . Rumus Flanagan adalah sebagai berikut:

$$r_{11} = 2 \left(1 - \frac{SB^2_{\text{gasal}} + SB^2_{\text{genap}}}{SB^2_{\text{total}}} \right)$$

Rumus ini lebih sederhana daripada rumus Spearman-Brown. Selain dengan rumus Spearman-Brown dan Flanagan, dapat pula dengan menggunakan rumus Rulon. Pada rumus Rulon, pertama ditentukan deviasi dari belahan skor gasal dan belahan skor genap. Langkah berikutnya mencari kuadrat simpangan

baku dari deviasi skor tersebut dan kuadrat simpangan baku dari skor total. Rumus Rulon adalah sebagai berikut:

$$R_{11} = 1 - \frac{SB^2 \text{ deviasi}}{SB^2 \text{ total}}$$

keterangan :

SB2deviasi = kuadrat simpangan baku skor deviasi

SB2total = kuadrat simpangan baku total

2. Faktor-Faktor Yang Mempengaruhi Reliabilitas Instrumen

Menurut Sukardi (2009) koefisien reliabilitas dapat dipengaruhi oleh waktu penyelenggaraan tes-retes. Interval penyelenggaraan yang terlalu dekat atau terlalu jauh, akan mempengaruhi koefisien reliabilitas.

Faktor-faktor lain yang juga mempengaruhi reliabilitas instrument evaluasi diantaranya sebagai berikut :

- a. Panjang tes, semakin panjang suatu tes evaluasi, semakin banyak jumlah item materi pembelajaran diukur.
- b. Penyebaran skor, koefisien reliabilitas secara langsung dipengaruhi oleh bentuk sebaran skor dalam kelompok siswa yang di ukur. Semakin tinggi sebaran, semakin tinggi estimasi koefisien reliabel.
- c. Kesulitan tes, tes normatif yang terlalu mudah atau terlalu sulit untuk siswa, cenderung menghasilkan skor reliabilitas rendah.
- d. Objektifitas, yang dimaksud dengan objektif yaitu derajat dimana siswa dengan kompetensi sama, mencapai hasil yang sama.

E. Beberapa Kesalahan dalam Pengukuran

Pengukuran adalah pemberian angka yang merefleksikan karakteristik suatu objek. Sebuah pengukuran belum tentu memberikan nilai atau angka sebenarnya atas suatu karakteristik yang diukur, melainkan lebih cenderung hanya memberikan nilai atau angka observasi atas karakteristik yang diukur. Contohnya, jika kita mengukur luas lautan yang ada di bumi, seberapa luaskah laut itu sebenarnya? Pengukuran luas lautan memberikan suatu angka yang merupakan nilai berdasarkan hasil observasi. Misalkan

hasil pengukuran tersebut menghasilkan angka 30 juta km persegi. Apakah angka ini merupakan angka sebenarnya atas luas lautan di bumi ini? Bisa jadi luas lautan sebenarnya adalah 35 juta km persegi. Dapat pula 28 juta km persegi. Selisih antara hasil observasi yang kita dapatkan dengan angka sebenarnya disebut error. Setiap pengukuran pasti mengandung unsur error (measurement error), baik itu yang sifatnya konstan maupun yang tidak konstan.

Menurut *True Score Model* skor suatu objek yang diukur (X_0) dipengaruhi oleh systematic error (X_S), random error (X_R), dan true score dari karakteristik objek yang diukur itu sendiri (X_T). Dalam persamaan matematis, ini dapat dilambangkan sebagai berikut:

$$X_0 = X_T + X_S + X_R$$

Systematic error (X_S) mempengaruhi akurasi pengukuran secara konstan (terus-menerus). Error ini mempengaruhi observed score (X_0) dengan sifat yang sama setiap kali pengukuran dilakukan, misalnya karena factor mekanis, method error dan systematic respondent error. Contoh method error adalah instrument alat ukur yang tidak bekerja dengan akurat sebagaimana mestinya. Misalnya, kita menimbang berat badan petinju kelas berat dunia Mike Tyson, maka alat timbangan yang kurang sempurna dapat membuat berat badan Tyson 500 gram kurang atau lebih dari berat badan sebenarnya. Method error juga dipengaruhi oleh factor mekanis seperti instruksi yang tidak jelas kepada responden pada kuisioner, tampilan kuisioner yang tidak sempurna sehingga tidak terbaca dengan jelas oleh responden, susunan pernyataan yang terlalu padat sehingga membingungkan responden, atau kurang jelasnya cara untuk menjawab skala yang diberikan.

Contoh systematic respondent error adalah adanya social desirability bias, yaitu kecenderungan seseorang untuk menjawab pertanyaan sedemikian rupa sehingga membuat dirinya kelihatan positif sesuai dengan norma yang standar yang diakui banyak orang. Misalnya, seseorang yang memiliki social desirability bias yang tinggi akan menjawab bahwa ia sangat puas dengan pekerjaannya (*over-reporting*), berkomitmen tinggi terhadap perusahaan tempat ia bekerja dan berkeinginan kecil sekali (*under-reporting*) untuk mencari pekerjaan baru di tempat lain. Contoh

lain untuk systematic respondent error adalah pengaruh factor inteligensia atau pendidikan responden. Inteligensia dan pendidikan responden bisa mempengaruhi akurasi hasil penelitian, terutama pada penelitian ilmu social. Misalkan si Dadi melakukan penelitian tentang sikap masyarakat di pesisir pantai terhadap pentingnya kelestarian terumbu karang. Tingkat pendidikan akan mempengaruhi cara berfikir mereka tentang pentingnya kelestarian terumbu karang terhadap produktivitas mereka sebagai nelayan.

Contoh systematic respondent error lainnya adalah acquiescence response sets, yaitu kecenderungan responden untuk setuju atau tidak setuju dengan item-item pada kuisisioner tanpa memandang apa pun isi item yang ditanyakan pada kuisisioner tersebut. ini adalah fenomena responden untuk menjawab “ya” atau “tidak” sesukanya. Dengan demikian, systematic error sifatnya melekat (inherent) dalam suatu pengukuran, baik karena error pada metode atau alat yang digunakan untuk pengukuran maupun karena error pada responden yang sifatnya permanen.

Random error (XR) mempengaruhi akurasi pengukuran karena factor transien atau factor situasional. Factor transien, seperti perubahan emosi, fisik, atau kelelahan responden bisa membuat jawaban yang diberikan responden kepada peneliti menjadi kurang atau tidak akurat. Factor situasional misalnya karena kehadiran orang lain, suara berisik atau gangguan lainnya sehingga mempengaruhi akurasi jawaban responden.

Error yang kita bahas di atas baru merupakan sebagian dari kemungkinan kesalahan yang ada. Masih ada sumber error lainnya, diantaranya non response error, yaitu kesalahan yang diakibatkan adanya beberapa responden yang termasuk dalam sampel tetapi tidak merespon penelitian. Penyebabnya bisa dua macam, yaitu menolak menjadi responden (refusals) dan sedang tidak ada ditempat (not-at-homes). Penolakan calon responden disebabkan oleh berbagai alasan, diantaranya: tidak memiliki waktu, tidak ingin diganggu, kuisisioner terlampaui panjang dan kompleks, topic penelitian tidak menarik, topic penelitian merupakan isu sensitive, sikap pewawancara kurang sopan, dan beraneka alasan lainnya. Selain itu ada omitted variable error, yaitu kesalahan yang dikarenakan adanya variable yang hilang, baik sengaja maupun tidak disengaja. Tipe

kesalahan berikutnya adalah sampling error, yaitu ketidakakuratan karena penentuan ukuran sampel yang salah, sampel yang dipilih tidak representative, penentuan sampling frame yang keliru dan sebagainya.