

PENGUKURAN DALAM PENELITIAN PSIKOLOGI

Rosleny Marliani

Fakultas Psikologi UIN Sunan Gunung Djati Bandung
Jl. Raya Cipadung No. 105 Bandung 10614 Telp (022) 7800525 email: rose_marlia@yahoo.com

Psychology is the science of human and animal behavior. It is also intended to get a clear description of any evidence or behavior of any population. For example, we want to know the correlation of teenagers' intelligence and their creativity. It is an example of such case that can be described and measured by using some indicators.

Measurement in psychology is very important because it is not only an art but also a science. As science it is empirical. It is of course relies on experiment and observation, rather than argument, opinion or belief.

Moreover, Morgan (1961) said that the progress of in science often depends upon the development of quantitative methods. Without such methods, science is limited to crude observation and classification. By using them we are able to refine data-based conclusions.

There are so many kinds of measurement that can be used in psychological field, such as test, personality inventory, attitude scale, etc. There are some important aspect that should be noted in measurement, such as reliability and validity.

Keywords: measurement, reliability, validity

Pendahuluan

Salah satu dari tujuan penelitian dalam bidang psikologi adalah untuk memperoleh pemahaman yang lebih luas tentang hubungan antara gejala, karakteristik, atau variabel dari suatu kelompok atau populasi tertentu. Sebagai contoh, kita mungkin ingin mengetahui ada tidaknya hubungan antara inteligensi dan kreatifitas pada kelompok siswa remaja madya atau pada kelompok individu yang berusia antara 15 hingga 18 tahun. Kita tak dapat mengamati secara langsung baik inteligensi maupun kreatifitas. demikian pula kita dapat melakukan pengamatan terhadap semua individu yang berusia antara 15 hingga 18 tahun. Meskipun demikian, itu tidak berarti bahwa kita tak dapat memperoleh jawaban tentang ada tidaknya hubungan antara inteligensi dan kreativitas, maupun tentang hubungan antara gejala atau

variabel-variabel yang lain. Kita dapat menggunakan indikator-indikator untuk mendefinisikan konstruk inteligensi atau kreatifitas. Indikator-indikator tersebut merupakan cara pendekatan dengan menggunakan contoh-contoh perilaku aktual yang dapat diamati secara langsung yang dapat diterima sebagai suatu indeks yang valid dari konstruk psikologis, seperti inteligensi, kreatifitas, motivasi, prestasi, sikap, minat, dsb.

Penggunaan indikator guna mendefinisikan suatu konstruk psikologis merupakan salah aspek pengukuran dalam penelitian. Pengukuran itu sendiri merupakan aspek penting dalam penelitian. Indikator merupakan jabaran operasional dari suatu konstruk psikologis yang hendak diukur dan diteliti. Indikator juga merupakan acuan darimana butir-butir alat ukur dikembangkan atau disusun.

Setelah memilih suatu konstruk yang hendak diteliti dan kemudian menjabarkannya ke dalam indikator-indikator dan butir-butir pengukuran, langkah selanjutnya yang harus dilakukan oleh peneliti adalah memilih atau menetapkan alat ukur atau instrumen penelitian. Dalam penelitian, istilah alat ukur atau instrumen penelitian digunakan untuk menunjuk pada berbagai macam alat pengumpul data. Beberapa jenis alat ukur tersebut bisa berupa tes atau alat ukur kepribadian (*inventory*) terstandar atau yang dikembangkan sendiri oleh peneliti, pedoman observasi, atau alat perekam data yang sudah tersedia (Gay, 1987).

Di samping mendefinisikan konstruk psikologis yang akan diukur dan menjabarkannya ke dalam indikator-indikator pengukuran, aspek penting yang lain dalam pengukuran adalah menetapkan jenis ukuran (data hasil pengukuran) yang akan dikumpulkan, apakah kuantitatif atau kualitatif. Jika data penelitian akan diekspresikan dalam bentuk kuantitas, misalnya, perlu dipikirkan juga apakah data kuantitas tersebut akan diekspresikan dalam bentuk skala nominal, ordinal, interval, atau rasio. Tulisan ini akan memberikan kajian secara ringkas berkenaan dengan aspek-aspek pengukuran tersebut.

Pembahasan

Hakikat Pengukuran

Pengukuran (*measurement*) merupakan suatu cara yang sistematis untuk menetapkan secara pasti bilangan-bilangan (angka-angka) atau nama-nama terhadap suatu obyek dan atribut-atributnya (Baltes, Reese & Nessel-

roade, 1988; Elmes, Kantowitz, & Roedriger III, 1992; Nunnally, 1978). Pengukuran merupakan aspek penting dalam penelitian psikologis maupun penelitian-penelitian dalam bidang lainnya. Seperti dikemukakan oleh Boring, seorang tokoh psikologi yang terkenal (Elmes, Kantowitz, & Roedriger III),

“...a science grows and progresses to the extent that it uses measurement ...testable hypotheses imply that the predicted outcomes can be measured....empirical observations in psychology are based on the measurement of dependent variables. Measurement then is an important aspect of scientific psychology (1992:75).

Baltes, Reese, & Nesselroade (1988) juga menyatakan bahwa pengukuran merupakan salah satu landasan (*cornerstone*) dalam penelitian empiris dari berbagai disiplin ilmiah. Pengukuran secara langsung merepresentasikan suatu cara yang dipilih oleh peneliti untuk mendefinisikan konsep-konsep penting. Dikemukakan lebih lanjut oleh Baltes, Reese, & Nesselroad, melalui pengukuran kita dapat menggambarkan dan merangkumkan variabel-variabel ke dalam angka, nama, atau indeks-indeks yang lain sehingga mudah untuk mengkomunikasikan kepada orang lain. Hasil pengukuran (misalnya angka) juga memberikan material kasar bagi kepentingan analisis statistik dan matematika untuk maksud menetapkan hubungan antara variabel-variabel yang diteliti.

Pengukuran suatu konstruk psikologis (misalnya pengukuran motivasi, sikap, atau konsep diri), bagaimanapun, relatif lebih rumit jika dibandingkan dengan pengukuran dalam

bidang eksakta atau fisik. Sebab, berbagai atribut dalam bidang fisik, seperti tinggi badan, berat badan, panjang rambut, dsb. dapat diukur secara langsung dalam arti jarak fisik, di samping aturan untuk menetapkan angka-angka yang mewakili atribut-atribut tersebut relatif lebih tegas dan sederhana. Sedangkan atribut-atribut karakteristik psikologis pada umumnya bersifat abstrak dan aturan pengukurannya relatif kurang tegas. Pengukuran beberapa konsep psikologis seperti ketergantungan, kekuatan ego, ekstroversi, kemandirian, dsb. Merupakan suatu konsep-konsep yang tidak bisa diukur secara langsung, tetapi harus diterjemahkan lebih dahulu ke dalam bentuk perilaku-perilaku khusus yang dapat diamati. Dalam hal ini, terdapat variasi yang luas dalam bagaimana peneliti mendefinisikan konsep-konsep tersebut tergantung pada kerangka pikir atau landasan teoretis yang digunakannya. Beberapa konsep psikologis juga tampak lebih rumit. Misalnya, konsep *attachment*, merupakan salah satu contoh konsep psikologis yang tidak hanya abstrak tetapi juga rumit karena melibatkan kombinasi antara organisme, seperti ibu dan anak.

Skala Pengukuran

Penelitian dalam bidang psikologi, demikian pula dalam bidang-bidang yang lainnya, selalu melibatkan isu tentang skala pengukuran (*types of measurement scales*). Penerapan skala pengukuran ini berkenaan dengan isu tentang interpretasi yang dapat diberikan terhadap bilangan-bilangan yang dihasilkan dari suatu pengukuran psikologis, dan legitimasi prosedur

matematis tertentu yang diterapkan oleh peneliti dalam suatu pengukuran psikologis. Misalnya, apakah suatu alat ukur inteligensi setara dengan meteran, dapatkah hasil pengukuran psikologis dianalisis secara sama seperti hasil pengukuran fisik, dsb. Dalam penelitian psikologis (juga dalam penelitian bidang lainnya), terdapat empat skala pengukuran, yakni: nominal, ordinal, interval, dan rasio (Ary, Jacob, & Razavieh, 1985; Baltes, Reese, & Nesselroad, 1988; Gay, 1987; McMillan & Schumacher, 2002). Berikut adalah penjelasan secara garis besar dari empat skala pengukuran tersebut.

1. Skala nominal

Skala nominal hanya memilahkan obyek atau atribut ke dalam kategori-kategori yang berbeda, seperti: jenis kelamin (laki-perempuan), tingkat pendidikan (dasar, menengah, tinggi), umur (0-6 tahun; 7 – 12 tahun; 13 – 18 tahun); jenis pekerjaan (pegawai negeri, wiraswasta, BUMN, dsb). Menurut Elmis, Kantowitz, & Rodrigger III (1992), skala nominal memiliki kelemahan dalam arti bahwa kita tak dapat melakukan banyak operasi matematis pada bilangan-bilangan nominal. Dicontohkan, ketika kita menempatkan orang ke dalam suatu kategori khusus, misalnya memberinya nama, nama tersebut hanya memberi informasi bahwa orang tersebut berbeda dari orang lain. Suatu bilangan nominal tidak membentuk kita dalam mengukur atribut-atribut orang tersebut secara lebih mendalam.

2. Skala ordinal

Skala ordinal mengukur perbedaan dalam besaran atau jarak (*magnitude*) suatu ob-

yek atau atribut. Tipe skala ini dapat kita peroleh jika kita menempatlan obyek atau atribut ke dalam suatu urutan ranking. Misalnya, kita menempatkan siswa-siswa ke dalam suatu urutan ranking atas dasar prestasi hasil belajarnya. Siswa pertama dari urutan ranking tersebut adalah siswa yang paling tinggi capaian prestasi akademik, siswa kedua adalah nomor dua dari atas dalam pencapaian prestasi belajar, dan seterusnya. Demikian pula kita dapat menempatkan seseorang ke dalam urutan ranking kekayaan, sikap prososial, tingkat motivasi, daya tarik, dsb. Jika kita membuat ranking suatu obyek atau atribut, satu hal yang perlu dicatat adalah bahwa perbedaan antara nilai skala yang berdekatan atau berbatasan tidak harus selalu sama untuk semua obyek atau atribut yang diranking. Sebagai contoh, siswa dalam ranking pertama dan kedua mungkin hampir berdekatan dalam tingkat capaian prestasi belajarnya, tetapi siswa kelima dan keenam mungkin berbeda agak jauh. Dengan kata lain, skala ordinal tidak memiliki interval yang sama dan dengan demikian kita juga tak dapat melakukan banyak operasi matematis terhadap jenis skala ini, misalnya dicari rata-ratanya atau simpangan bakunya. Dapat dikatakan bahwa skala ordinal merupakan : (1) pengurutan seperangkat obyek kedalam suatu "ranking" (dari paling atas hingga paling bawah, atau sebaliknya) menurut atribut tertentu; (2) tidak ada indikasi berkenaan dengan "seberapa banyak" suatu atribut dimiliki oleh obyek; dan (3) tidak ada indikasi mengenai jarak atau perbedaan atribut dari obyek yang diamati/diukur. Jadi, dalam contoh pengurutan

prestasi belajar siswa, tidak ada informasi menyangkut seberapa besar prestasi belajar dari siswa yang satu berbeda dengan siswa lainnya, kecuali informasi bahwa siswa pertama > siswa kedua > siswa ketiga > siswa keempat dan seterusnya.

3. Skala interval

Dalam skala interval terdapat perbedaan (interval) yang sama dan tegas menyangkut suatu atribut tertentu antara obyek yang satu dengan lainnya. Jadi dalam skala interval lebih banyak operasi matematis yang dapat dilakukan, seperti menambah, mengurangi, membagi, mengalikan, mencari rata-rata, atau menemukan simpangan bakunya. Dengan kata lain, skala interval memiliki sifat perbedaan, urutan, dan interval yang sama. Sebagai contoh, perbedaan skor IQ antara 100 dan 110 adalah sama dengan perbedaan skor IQ antara 120 dan 130. Demikian pula, perbedaan prestasi belajar dari 5-6 adalah sama dengan perbedaan prestasi belajar dari 8-9. Skor IQ dan prestasi belajar tersebut merupakan dua bentuk contoh dari skala interval.

4. Skala rasio

Skala rasio memiliki empat karakteristik pengukuran: perbedaan, jarak, interval sama, dan nol mutlak (Ary, Yacob, & Razavieh, 1985; Baltes, Reese, & Nesselroad, 1988; Gay, 1987; McMillan & Schumacher, 2002). Skala rasio dapat memberikan banyak informasi dan selalu dinilai sebagai bentuk skala yang sangat bermanfaat dalam pengukuran psikologis. Karena skala rasio memiliki nilai nol mutlak (*true zero*), maka skala ini memungkinkan kita untuk menetapkan rasio dari nilai-nilai skala

(oleh karena itu skala ini disebut skala rasio). Bahkan jika skor IQ diukur dalam skala interval, kita dapat mengatakan bahwa IQ 120 adalah dua kali lebih besar dari IQ 60, atau lebih besar dari IQ 100.

Jenis-Jenis Alat Ukur Psikologis

Terdapat beberapa jenis alat ukur psikologis yang biasanya digunakan dalam penelitian psikologis, antara lain: tes, inventori kepribadian, skala sikap, observasi, dan wawancara (Ary, Yacob, & Razavieh, 1985; Baltes, Reese, & Nesselroad, 1988; Gay, 1987; McMillan & Schumacher, 2002). Berikut adalah penjelasan secara garis besar tentang jenis-jenis alat ukur tersebut.

1. Tes

Menurut Ary, Jacobs, & Razavieh (1985), tes merupakan alat ukur yang sangat penting dalam penelitian psikologi. Tes merupakan seperangkat stimuli yang disajikan kepada individu untuk mendatangkan atau memperoleh respon-respon yang diekspresikan dalam bentuk skor angka. Skor ini didasarkan pada sampel perilaku individu yang representatif atau pada indikator-indikator dari atribut yang diukur oleh suatu tes. Dalam penelitian psikologi, dikenal adanya tes terstandar (tes baku) dan tes tak terstandar (disusun dan dikembangkan sendiri oleh peneliti guna mengukur atribut yang sedang diamati). Termasuk dalam tes terstandar antara lain adalah tes inteligensi (*intelligence test*), yaitu tes yang digunakan untuk mengukur tingkat kecerdasan subjek, dalam arti kemampuannya untuk mempersepsi hubungan, memecahkan masalah, dan

menerapkan pengetahuan yang dimiliki dalam berbagai macam konteks; tes prestasi (*achievement test*), yaitu tes yang digunakan untuk mengukur kemampuan skolastik subjek; dan tes kepribadian (*projective test*), yakni suatu tes yang digunakan untuk mengukur sifat-sifat atau karakteristik kepribadian subjek. Suatu tes dikatakan terstandar jika telah memiliki beberapa properti (karakteristik) tes baku seperti: standar norma, validitas dan reliabilitas, dan petunjuk dalam mengadministrasikan dan penskorannya (Friedenburg, 1995).

2. Inventori kepribadian (*personality inventory*)

Karakteristik kepribadian di samping dapat diukur melalui tes projektif, juga dapat diukur melalui tes non proyektif, yakni yang biasa di sebut dengan inventori kepribadian. Salah satu bentuk inventori ini adalah laporan (*self report*), dan salah satu bentuk laporan diri tersebut adalah angket (*questionnaire*). Dalam inventori, subjek disajikan sejumlah pernyataan yang menggambarkan pola-pola perilaku tertentu dan diminta untuk menyatakan apakah pola-pola perilaku yang dinyatakan tersebut merupakan karakteristik perilakunya atau bukan, dengan menjawab *ya* atau *tidak*, atau dengan memberikan cek pada salah satu pilihan jawaban yang disediakan. Seperti halnya tes, inventori ada yang terstandar dan tak terstandar. Beberapa contoh inventori terstandar antara lain adalah: *California F-Scale*, yang digunakan untuk mengukur *autoritarianisme*; dan *Cattell's Sixteen Personality Factor Questionnaire*, yang digunakan untuk mengukur sejumlah sifat. Beberapa inventori lain yang

banyak digunakan dalam penelitian antara lain adalah *Minnesota Multiphasic Personality Inventory*, *the Guilford-Zimmerman Temperament Survey*, *the Mooney Problem Check List*, dan *the Edwards Personal Schedule*. Inventori telah banyak digunakan dalam penelitian-penelitian untuk mengetahui hubungan antara karakteristik kepribadian dengan beberapa variabel seperti inteligensi, prestasi, sikap, *underachievement* dari beberapa kelompok populasi atau subjek tertentu (Ary, Yacobs, & Razavief, 1985).

3. Skala sikap (*attitude scale*)

Menurut beberapa penulis (Ary *at al.*, 1985; gay, 1987; Friedenburg, 1995), skala sikap merupakan suatu alat ukur yang digunakan untuk mengukur sikap, nilai, dan karakteristik lain. Dalam skala sikap berisikan nilai-nilai bilangan untuk menilai subjek, obyek, atau perilaku-perilaku untuk maksud mengkuantifikasikan atau mengukur kualitas-kualitas. Skala sikap berbeda dengan tes, sebab tidak seperti halnya hasil tes, hasil pengukuran skala sikap tidak menyatakan kekuatan atau kelemahan, keberhasilan atau kegagalan. Skala sikap mengukur seberapa jauh individu memiliki karakteristik nilai, keyakinan, minat, atau pandangan terhadap sesuatu. Sebagai contoh, skala sikap dapat digunakan untuk mengukur sikap remaja terhadap partai politik, pemilu, atau pengurusan untuk kepentingan pembangunan. Banyak peneliti mendefinisikan sikap sebagai afek (perasaan) positif atau negatif terhadap suatu kelompok, institusi, konsep, atau obyek sosial tertentu. Dengan kata lain, pengukuran sikap pada dasarnya adalah menempatkan indi-

vidu dalam suatu kontinum positif (*favourable*) – negatif (*unfavourable*) terhadap suatu obyek sikap.

Terdapat beberapa bentuk skala sikap yang dapat digunakan oleh peneliti sebagai acuan dalam mengembangkan skala sikap, yaitu: (1) *summated rating scales* (skala Likert); (2) *equal-appering intervals scales* (skala Thurstone); (3) *cumulative scales* (skala Guttman); dan *semantic differential scales* (Ary *at al.*, 1985; Gay, 1987; Friedenberg, 1995). Dari empat model skala sikap tersebut, skala Likert merupakan tipe yang paling banyak digunakan.

Skala Likert mengukur sikap subjek terhadap suatu obyek sikap (topik) dengan cara meminta subjek untuk menyatakan apakah ia sangat setuju, setuju, tidak tahu, tidak setuju, atau sangat tidak setuju. Topik atau obyek sikap tersebut disajikan melalui pernyataan-pernyataan yang diekspresikan dalam bentuk kalimat positif (*favorable*) dan kalimat negatif (*unfavorable*). Respon subjek selanjutnya diskor dengan cara memberikan bobot. Misalnya, untuk respon terhadap pernyataan positif, skor yang diberikan adalah 5, 4, 3, 2, dan 1 untuk pernyataan sangat setuju, setuju, tidak tahu, tidak setuju, dan sangat tidak setuju. Sedangkan untuk pernyataan negatif, pemberian skornya dilakukan dengan cara sebaliknya.

Jika skala Likert mengukur sikap dengan cara meminta subjek untuk menyatakan tingkat kesetujuan atau ketidak setujuannya terhadap suatu pernyataan, skala Thurstone mengukur sikap dengan cara menyajikan suatu pernyataan tentang suatu topik yang merentang dari sangat positif (*very favorable*), netral

(*neutral*), hingga sangat negatif (*very unfavorable*), dan meminta subjek untuk memilih dari pernyataan-pernyataan tersebut yang sangat sesuai dengan sikapnya.

Skala Guttman juga disebut sebagai teknik kumulatif. Skala ini dimaksudkan untuk mengatasi kelemahan yang ada pada skala Likert maupun Thurstone. Kritik yang ditujukan terhadap skala Likert maupun Thurstone adalah bahwa kedua skala tersebut berisikan pernyataan-pernyataan yang heterogen tentang berbagai macam dimensi obyek sikap (Ary *at al.*, 1985). Sebagai contoh, skala Thurstone mengukur sikap terhadap perang, tidak memisahkan pernyataan jika dari pernyataan-pernyataan yang menyangkut dampak ekonomi yang disebabkan oleh perang, atau merefleksikan kemungkinan aspek-aspek lain dari sikap terhadap perang. Oleh karena itu, skala tersebut tampak mengkombinasikan beberapa dimensi pada satu skala dan memunculkan problem dalam membuat interpretasi yang tegas dari skor yang diperoleh. Untuk mengatasi problem yang terdapat skala Likert dan Thurstone, Guttman mengembangkan suatu skala yang menekankan hanya pada satu dimensi (*unidimensional scale*), dan dimaksudkan untuk mengukur sikap subjek terhadap satu dimensi obyek. Suatu sikap dipandang sebagai *unidimensional* hanya jika sikap tersebut menghasilkan suatu skala kumulatif – yakni skala dimana butir-butirnya memiliki hubungan satu satu lain. Misalnya, jika subjek setuju dengan butir nomor 2, ia juga setuju dengan butir 1; subjek yang setuju dengan butir 3 juga setuju dengan butir 1 dan 2, dan

seterusnya. Subjek yang menyetujui butir-butir khusus dalam model skala ini akan memperlihatkan skor tinggi pada total skala dibandingkan subjek yang tidak menyetujui.

Skala diferensial semantik merupakan pendekatan lain dalam mengukur sikap. model skala ini dikembangkan oleh Osgood, Suci, dan Tannenbaum (Ary *at al.*, 1985; *Friedenberg*, 1995). Skala ini didasarkan pada asumsi bahwa obyek sikap memiliki dua bentuk makna yang berbeda bagi individu, yakni: denotatif (makna harfiah dari suatu kata) atau konotatif (makna yang ditunjuk). Dua makna tersebut dapat dinilai secara bebas (*independent*). Orang pada umumnya lebih mudah untuk menyatakan makna denotatif dari suatu obyek alih-alih makna konotatif. Meskipun demikian, menurut Ary *at al.*, (1985), kita dapat mengukur makna konotatif dari suatu obyek secara tidak langsung dengan cara meminta individu untuk menilai obyek dengan menggunakan sejumlah kata sifat bipolar. Berdasarkan pada studi analitik yang dilakukannya, Osgood *at al.* menemukan tiga kluster kata sifat: (1) *evaluatif*, berisikan beberapa kata sifat seperti baik atau buruk, berguna atau tak berguna; (2) *potensi*, meliputi kata-kata sifat seperti kuat atau lemah, berat atau ringan; dan (3) *aktivitas*, berkaitan dengan kata sifat seperti aktif atau pasif, cepat atau lambat. Di antara tiga kluster tersebut, dimensi evaluatif tampak paling signifikan untuk mengukur sikap.

4. Beberapa teknik pengukuran lain

Di samping berbagai macam alat ukur seperti telah dikemukakan, terdapat beberapa teknik lain yang sering digunakan dalam

penelitian psikologi, yakni: observasi dan wawancara.

Dalam banyak kasus, melakukan observasi atau pengamatan langsung terhadap perilaku individu merupakan metode pengukuran yang sangat krusial (Ary *at al.*, 1985; Elmes *at al.*, 1992). Dalam teknik observasi, peneliti mengidentifikasi perilaku-perilaku yang ingin dikaji dan menggunakan suatu prosedur yang sistematis untuk mengidentifikasi, mengkategorikan, dan merekam perilaku subjek dalam cara yang terencana dan netral. Secara khusus, metode observasi telah banyak digunakan untuk secara luas dalam penelitian yang melibatkan subjek bayi dan anak-anak pra sekolah. Menurut McMillan & Schumacher (2001), semua teknik pengumpulan data pada dasarnya melibatkan beberapa bentuk observasi. Dalam artian umum, kata observasi digunakan untuk menggambarkan data yang dikumpulkan, tanpa memperhatikan teknik yang digunakan dalam penelitian. Meskipun demikian, metode-metode penelitian observasional menunjuk pada mode khusus dalam mengumpulkan informasi dan yang sangat berbeda dengan metode lain seperti wawancara atau laporan diri (angket). Sebagai suatu teknik pengumpulan data, metode observasi sangat menyandarkan pada kepekaan pengamatan dan pendengaran peneliti dalam merekam data yang diinginkan.

Observasi dapat dilakukan secara partisipatif dan non partisipatif. Jika observasi dilakukan secara partisipatif, peneliti ikut terlibat bersama subjek dalam suatu aktivitas yang sedang diamati. sebagai contoh, jika peneliti ingin mengumpulkan data tentang tingkat ke-

terlibatan subjek dalam memecahkan suatu masalah dalam kelompok, peneliti ikut terlibat dalam kegiatan tersebut, misalnya dengan menjadi anggota kelompok. Jika observasi dilakukan dengan cara non partisipatif, peneliti hanya mengamati dari jarak tertentu terhadap perilaku subjek. Cara ini tampak misalnya dalam suatu pengamatan yang menggunakan *one way screen* dalam suatu laboratorium.

Untuk dapat melakukan observasi dengan cermat dan teliti, peneliti perlu merumuskan aspek-aspek apa saja yang akan diamati ke dalam suatu alat yang disebut dengan pedoman observasi. Terdapat dua bentuk pedoman observasi, yakni daftar cek (*check list*) dan skala penilaian (*rating scale*). Jika peneliti menggunakan daftar cek, maka ia mengamati munculnya aspek-aspek perilaku subjek sebagaimana telah tercantum dalam pedoman dan kemudian memberikan cek pada aspek perilaku tersebut. Jika peneliti menggunakan skala penilaian, hanya ia tidak hanya merekam munculnya perilaku yang diamati tetapi juga frekuensi (seberapa sering) dari munculnya perilaku tersebut.

Wawancara pada dasarnya merupakan suatu metode pengumpul data yang tidak berbeda dengan metode laporan diri atau angket, tetapi tidak seperti halnya dalam metode angket, metode wawancara melibatkan komunikasi langsung secara lisan antara peneliti dengan subjek sumber data atau responden (McMillan & Schumacher, 2001). Karena dilakukan melalui komunikasi langsung, maka metode wawancara memiliki kelebihan dibanding metode lain, dalam arti bahwa wawancara da-

pat mengungkap data secara lebih mendalam tentang aspek-aspek perilaku subjek. Demikian pula – bagi para peneliti yang mahir – melalui wawancara akan dapat diketahui apakah subjek memberikan jawaban dengan jujur atau tidak jujur.

Agar wawancara berjalan lebih terarah dan sistematis, seperti halnya dalam teknik observasi, dalam teknik wawancara peneliti perlu untuk menggunakan suatu pedoman wawancara. Dalam pedoman wawancara tersebut berisikan pokok aspek perilaku yang akan dieksplorasi. Pengembangan suatu pedoman wawancara tidak berbeda dengan pengembangan alat ukur lainnya, yaitu: menetapkan tujuan wawancara, menuliskan pertanyaan, menetapkan format umum dan khusus, dan uji coba (McMillan & Schumacher, 2001). Dibandingkan dengan metode observasi dan angket, metode wawancara lebih bersifat adaptif dan fleksibel. Artinya, metode ini dapat digunakan untuk berbagai macam problem yang berbeda dan tipe orang yang berbeda. Peneliti juga dapat mengelaborasi atau membuat klarifikasi lebih jauh terhadap respon-respon subjek, sehingga peneliti dapat memperoleh data yang lebih lengkap dan lebih mendalam. Peneliti juga dapat merekam perilaku non verbal dari subjek ketika sedang diwawancarai.

Meskipun wawancara memiliki beberapa kelebihan, ia juga memiliki beberapa kelemahan, antara lain: peneliti sulit menghindari diri dari subjektivitasnya dalam memberikan penilaian terhadap respon subjek; mengkonsumsi banyak waktu dan tenaga, khususnya jika penelitian dilakukan terhadap sejumlah

besar subjek; dan kurang anonim karena hampir semua identitas subjek dapat dikenali oleh peneliti. di samping itu, wawancara tidak dapat dilaksanakan oleh semua orang, kecuali oleh mereka yang telah terlatih dan memiliki keahlian khusus.

Isu Tentang Validitas Dan Reliabilitas Alat Ukur

Validitas (*validity*) dan reliabilitas (*reliability*) merupakan salah satu karakteristik (*property*) penting dalam setiap alat ukur psikologis (Ary *at al.*, 1985; Baltes *at al.*, 1988; Elmes *at al.*, 1992; Fridenberg, 1995). Menurut beberapa penulis tersebut, validitas menunjuk pada seberapa jauh suatu alat ukur mengukur apa yang hendak diukur. Sedangkan reliabilitas menunjuk pada seberapa jauh suatu alat ukur memberikan hasil yang konsisten tentang apa yang diukur.

1. Validitas

Baltes *at al.* (1988) dan Friedenber (1995) menyebutkan adanya empat bentuk validitas, yakni validitas muka (*face validity*), validitas isi (*content validity*), validitas empiris (*emperical validity*), dan validitas konstruk (*construct validity*).

a. *Validitas muka*. Validitas muka menunjuk menyatakan bahwa suatu alat ukur tampak mengukur apa yang ingin diukur. Suatu tes aritmatik, misalnya, kurang memiliki validitas muka sebagai tes kosa kata, dan tes kosa kata kata kurang memiliki validitas muka sebagai tes aritmetik. Tes aritmetik tampak memiliki validitas muka jika untuk mengukur kemampuan memecahkan problem aritmetik, demikian pula tes kosa kata tampak memiliki validitas muka jika ia

digunakan untuk mengukur penguasaan dalam kosa kata.

- b. *Validitas isi*. Suatu alat ukur dikatakan memiliki validitas isi jika alat ukur tersebut berisikan butir-butir yang dapat mewakili domain dari atribut yang diteliti atau diukur. Sebagai contoh, suatu tes kosa kata dikatakan memiliki validitas isi berisikan butir-butir tes yang mengandung semua kosa kata dalam bahasa yang diteliti.
- c. *Validitas empirik*. Validitas empirik memiliki dua macam bentuk, yakni: prediktif dan konkuren (*concurrent*). Keduanya berhubungan dengan hubungan empirik antara skor hasil pengukuran dengan suatu kriteria (kriteria) tetapi berbeda dalam waktu pengumpulan data kriteria (Ary *at al.*, 1985). Validitas prediktif menunjuk pada seberapa jauh skor-skor pada suatu alat ukur dapat digunakan untuk memprediksikan skor-skor pada alat ukur yang lain. Validitas ini dapat ditetapkan dengan cara mengkorelasikan dua perangkat skor dari dua perangkat tes. Suatu alat ukur dapat memiliki beberapa macam validitas prediktif; artinya, suatu alat ukur mungkin sangat valid sebagai prediktor dalam beberapa situasi, agak valid sebagai prediktor dalam situasi yang lain; dan kurang valid pada situasi yang lain lagi. Validitas konkuren berbeda dari validitas prediktif hanya dalam waktu pengukuran. Dalam validitas konkuren, suatu alat ukur diberikan secara simultan, dan dalam validitas prediktif suatu alat ukur yang dipre-

diksikan diberikan setelah alat ukur prediktor

- d. *Validitas konstruk*. suatu alat ukur dikatakan memiliki validitas konstruk jika berisikan butir-butir yang merupakan jabaran operasional dari domain atau karakteristik psikologis yang diukur, atau hasil yang diperoleh adalah konsisten dengan yang dipreskripsikan oleh teori yang digunakan sebagai acuan. Sebagai contoh, suatu teori menyatakan bahwa terdapat suatu hubungan antara beberapa variabel dan beberapa kriteria, misalnya antara inteligensi dan kecepatan dalam belajar. Jika suatu alat ukur dimaksudkan untuk mengukur inteligensi, alat ukur tersebut dikatakan memiliki validitas konstruk jika hasil pengukuran dari alat ukur tersebut dapat memprediksikan kecepatan dalam belajar. Validitas konstruk pada dasarnya merupakan validitas empirik, tetapi tidak seperti halnya validitas prediktif dan validitas konkuren, validitas konstruk lebih didasarkan pada suatu prediksi yang dinyatakan oleh suatu teori.

Validitas konstruk merupakan tipe validitas yang banyak dihitung dalam pengembangan alat ukur psikologis. Sebagaimana dikemukakan oleh Friedenberg (1995), pemeriksaan (studi) tentang validitas alat ukur psikologis (konstruk psikologi) mayoritas memusatkan perhatian pada validitas konstruk. Salah satu pendekatan yang paling umum digunakan untuk menghitung validitas konstruk adalah dengan mengkorelasikan skor pada setiap butir suatu alat ukur dengan skor total (*item-total correlation*) (Frieden-

berg, 1995). Koefisien korelasi antara skor butir dan skor total dapat menunjukkan kemampuan dari suatu butir alat ukur untuk memprediksikan skor total butir, atau derajat kesesuaian/keterpaduan antara suatu butir dengan perangkat butir yang lain dalam keseluruhan alat ukur (Friedenberg, 1995). Landasan berpikir dari pendekatan ini adalah: jika butir alat ukur dan keseluruhan butir alat ukur mengukur atribut yang sama, maka skor pada suatu butir dapat dikorelasikan dengan skor total. Jika suatu konstruk psikologis terdiri atas beberapa faktor atau atribut, menurut Friedenberg (1995), untuk menghindari terjadinya bias dalam perolehan koefisien korelasi, maka dalam prosedur ini peneliti sebaiknya tidak mengkorelasikan skor butir dengan skor total butir dari seperangkat alat ukur, tetapi mengkorelasikan skor butir dengan skor total dari masing-masing atribut (korelasi butir-total faktor/aspek).

2. Reliabilitas

a. Pengertian

Dalam artian yang sangat luas, reliabilitas menunjuk pada konsistensi suatu alat ukur; artinya, jika suatu alat ukur digunakan untuk mengukur aspek atau atribut yang sama seharusnya hasil yang diperoleh relatif tidak berbeda (Baltes *at al.*, 1988). Baltes *at al.* juga mengutip pernyataan Anastasi (1968) bahwa, sekurang-kurangnya dalam domain pengukuran psikologis, konsistensi merupakan esensi dari reliabilitas. Dalam pengukuran, peneliti seharusnya telah yakin bahwa nilai-nilai bila-

ngan telah diberikan untuk peristiwa-peristiwa atau atribut yang diamati dalam cara yang akurat, persis, dan konsisten. Meskipun demikian, menurut Baltes *at al.*, selalu terdapat hambatan untuk mencapai kriteria ideal tersebut, dan hambatan tersebut merupakan sumber kesalahan pengukuran.

Dalam konteks pengukuran psikologis, para penulis seperti Anastasi (1968) dan Nunnally (1967, 1970) telah mendaftar secara eksplisit berbagai macam pengaruh yang dapat memperendah reliabilitas suatu alat ukur dalam lingkungan khusus. Nunnally (1970), misalnya, menyebutkan faktor-faktor seperti kurangnya instruksi yang jelas dan rinci dalam tes terstandar, kesalahan dalam penskoran, kesalahan yang berhubungan dengan pengaruh subjektivitas, lingkungan tempat dilakukannya pengukuran, cara pengambilan sampel, dan fluktuasi individu sebagai faktor yang dapat memperendah reliabilitas alat ukur.

b. Menetapkan reliabilitas

Reliabilitas suatu alat ukur dapat ditetapkan dengan cara menggunakan indeks reliabilitas. Indeks reliabilitas ini dapat diperoleh melalui tiga pendekatan, yakni: *test-retest*, *equivalent-form* atau *alternate-form*, dan *split-half* (Friedenberg, 1995). Dalam teknik pertama, suatu alat ukur diadministrasikan dua kali kepada sekelompok subjek yang sama pada waktu yang berbeda. Ini akan menghasilkan dua perangkat skor. Indeks reliabilitas dapat ditemukan dengan cara mengkorelasikan dua perangkat skor tersebut.

Dalam teknik kedua, *equivalent-forms* atau sering disebut juga dengan istilah *alter-*

nate-form, digunakan jika terdapat kemungkinan bahwa subjek akan mengingat responnya pada butir-butir alat ukur. Dalam teknik ini, peneliti tidak mengkorelasikan dua perangkat skor yang dihasilkan dari dua kali pengukuran dengan menggunakan alat ukur dan kelompok subjek yang sama, tetapi peneliti mengkorelasikan hasil pengukuran dengan hasil pengukuran dari alat ukur yang lain yang dinilai ekuivalen (sama). Jadi, disamping mengadministrasikan alat ukur yang akan dilihat reliabilitasnya, peneliti juga mengadministrasikan alat ukur lain yang ekuivalen dan telah dinyatakan valid dan reliabel. Dua perangkat skor hasil pengukuran tersebut kemudian dikorelasikan.

Dalam teknik ketiga, peneliti hanya mengadministrasikan satu alat ukur kepada sekelompok individu kemudian membelah alat ukur tersebut ke dalam dua bagian sama besar. Pembagian ini bisa dilakukan atas dasar nomor butir, misalnya ganjil atau genap atau dengan membaginya secara acak. Indeks reliabilitas dapat ditemukan dengan cara mengkorelasikan perangkat skor pada belahan yang satu dengan perangkat skor pada belahan yang lain. Teknik ketiga ini merupakan pendekatan yang paling banyak digunakan. Terdapat beberapa formula terkenal untuk menghitung indeks reliabilitas dari teknik ketiga ini, antara lain adalah Kuder, Spearman Brown, dan Alpha Cronbach, tetapi formula Alpha Cronbach dipandang paling baik karena didasarkan pada varian skor butir (Friedenberg, 1995).

Reliabilitas dari suatu alat ukur selanjutnya didasarkan apakah indeks reliabilitasnya – koefisien korelasi yang ditemukan –

signifikan atau tidak. Jika koefisien korelasi yang ditemukan signifikan, maka alat ukur dikatakan reliabel, dan jika tidak signifikan, alat ukur dikatakan tidak reliabel. Signifikansi dari koefisien korelasi ditetapkan secara statistik, yakni dengan cara membandingkan nilai koefisien korelasi yang ditemukan dengan harga kritiknya. Harga kritik tersebut dapat ditemukan pada tabel-tabel statistik pada level taraf signifikansi (misalnya 99% atau 95%) atau taraf probabilitas (misalnya .01 atau .05) yang diinginkan, dan pada derajat kebebasan yang terhitung (ditetapkan berdasarkan jumlah sampel).

Simpulan

Salah satu tugas penting bagi peneliti dalam bidang psikologi adalah memilih instrumen pengumpul data atau alat ukur yang memadai untuk tujuan mengkuantifikasikan informasi atau data penelitian. Dalam mengkuantifikasikan data penelitian, peneliti dapat memilih untuk mengekspresikannya kedalam bentuk skala nominal, ordinal, interval, atau rasio. Untuk mengumpulkan data penelitian, peneliti dapat memilih alat ukur psikologi dalam bentuk tes, inventori, skala sikap, atau teknik pengumpul data lain seperti observasi atau wawancara. Peneliti dapat menggunakan alat ukur yang telah tersedia atau mengembangkannya sendiri.

Salah satu aspek penting yang tidak bisa dipisahkan dari suatu alat ukur adalah validitas dan reliabilitas dari alat ukur yang dipilih atau digunakan dalam penelitian. Validitas menunjuk pada tingkat seberapa jauh suatu alat ukur

mengukur apa yang hendak atau seharusnya diukur, sedangkan reliabilitas menunjuk pada seberapa jauh suatu alat ukur menghasilkan hasil pengukuran yang konsisten. Terdapat empat macam validitas, yakni: validitas muka, validitas isi, validitas empirik, dan validitas konstruk. Validitas konstruk merupakan macam validitas yang banyak diperhatikan. Salah satu teknik untuk menetapkan validitas konstruk adalah dengan mengkorelasikan skor butir dengan skor total, atau antara skor butir dengan skor total. Reliabilitas suatu alat ukur dapat ditetapkan sekurang-kurangnya melalui tiga teknik, yakni: *test-retest*, *equivalent-form*, dan *split-half*.

Daftar Pustaka

- Ary, D., Jacobs, L.C., & Razavieh, A. (1985). *Introduction to Research in Education*. 3rd. ed. New York: Holt, Rinehart and Winston.
- Baltees, P.B., Reese, H.W., & Nesselroads, J.R. (1988). *Introduction to research Methods in Life-span Developmental Psychology*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Elmes, D.G., Kantowitz, B.H., & Roedriger III, H.L. (1992). *Research methods in Psychology*. 4th. ed. New York: West publishing Company.
- Friedenberg, Lisa. (1995). *Psychological Testing. Design, Analysis, and Use*. Boston: Allyn and Bacon.
- Gay, L.R. (1987). *Educational Research. Competencies for Analysis and Application*. 3rd. ed. Columbus: Merrill Publishing Company.
- McMillan, J.H., & Schumacher, S. (2001). *Research in Education. A Conceptual Introduction*. 5th. ed. New York: Longman.

