

Validitas Eksperimen

T. Dicky Hastjarjo¹

Fakultas Psikologi
Universitas Gadjah Mada

Abstract

The students have understood the definition of instrument's validity. this article will explain the definition of experiment's validity that is different with instrument's validity. it also will describe the four types of experiment's validity and the threats againts them.

Pengertian validitas (*validity*) telah kita kenal lewat tulisan Hadi (1977), Masrun (1979), Azwar (1997) serta Suryabrata (2000). Pengertian validitas pada keempat tulisan tersebut mengacu pada pengertian yang dihasilkan oleh komisi gabungan antara APA (*American Psychological Association*), AERA (*American Educational Research Association*) dan *the National Council on Measurement Used in Education* yang membagi tiga macam validitas, yakni (a) Validitas isi (*content validity*), (b) Validitas konstruk (*construct validity*) serta (c) Validitas berdasarkan kriterium (*criterion-related validity*) (Kerlinger, 1986; Kerlinger & Lee, 2000). Ketiga macam validitas tadi berkaitan dengan masalah pengukuran (*measurement*) yang inti pokoknya menyangkut sejauhmana sebuah alat ukur psikologis dapat dipercaya mengukur apa yang seharusnya diukur (Suryabrata, 2000). Uraian berikut akan membahas validitas dalam bidang lain, yaitu validitas eksperimen.

Pengertian validitas eksperimen

Seperti diketahui bersama sebuah eksperimen psikologi ditandai dengan ada-

nya manipulasi variabel independen atau dinamakan juga perlakuan (*treatment*). Peneliti akan memunculkan variasi variabel independen atau disebut kondisi perlakuan (*treatment condition*) untuk dikaji pengaruhnya terhadap variabel dependen. Peneliti kemudian melakukan analisis data dengan statistik dan membuat kesimpulan/inferensi berdasar hasil analisis statistik tersebut. Validitas eksperimen menunjuk pada ancar-ancar kebenaran sebuah inferensi (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002), sebab ketika kita menyatakan sesuatu adalah valid maka kita membuat penilaian mengenai sejauh mana bukti yang relevan mendukung kebenaran atau kesalahan inferensi tentang sesuatu itu (Shadish dkk, 2002, hal. 34). Misalnya, Affandi dan Hastjarjo (2010) memanipulasi tipe *goal setting* menjadi tiga macam (*learning goal*, *performance goal*, dan *do your best goal*) dan meneliti pengaruhnya terhadap prestasi bahasa Inggris siswa SMU kelas XII. Tiga kelas XII yang berbeda masing-masing mendapat tipe *goal setting* yang berbeda pula, lalu prestasi bahasa Inggris mereka diukur. Efikasi diri dan kemampuan awal bahasa Inggris menjadi kovariabel sehingga analisis kovarians menjadi teknik statistik untuk menganalisis data. Mereka menyimpulkan bahwa prestasi bahasa

¹ Korespondensi mengenai isi artikel ini dapat melalui: dickyh@ugm.ac.id atau tdhastjarjo@gmail.com

Inggris siswa SMU kelas XII pada kelompok *learning goal* lebih tinggi daripada kelompok *performance goal* dan *do your best goal* dengan efikasi diri dan kemampuan awal bahasa Inggris sebagai kovariabel. Konsep validitas eksperimen akan mempermasalahkan kebenaran kesimpulan yang telah dibuat Affandi dan Hastjarjo tersebut. Validitas adalah ancar-ancar mengenai kebenaran sebuah inferensi sebab dalam sebuah eksperimen psikologi kebenaran mutlak sebuah kesimpulan tidak ada sehingga penggunaan istilah valid atau benar dan invalid atau salah senantiasa dipahami dengan memakai tambahan kata "tentatif" atau "bersifat kira-kira". Meskipun demikian penggunaan sehari-hari dalam penelitian eksperimen kata tambahan itu dihilangkan.

Shadish, dkk (2002) menegaskan bahwa validitas merupakan properti inferensi, bukan properti rancangan eksperimen (*design*) atau metode oleh karena rancangan yang sama mungkin akan memiliki inferensi yang lebih valid atau kurang valid tergantung situasinya. Misalnya, sebuah rancangan eksperimen acak (*randomized experiment*) tidak menjamin peneliti untuk membuat inferensi yang valid mengenai adanya hubungan kausal deskriptif. Sebab mungkin saja dalam eksperimen acak tersebut faktor atrisi subjek (jumlah subjek yang mengundurkan diri untuk mengikuti eksperimen sampai selesai) yang berbeda antara kelompok eksperimen dan kontrol akan melemahkan proses randomisasi, *power* mungkin terlalu lemah untuk mendeteksi adanya efek, atau teknik statistik yang tidak cocok untuk menganalisa data. Jadi adalah keliru menyatakan bahwa sebuah eksperimen acak memiliki validitas internal atau bersifat valid secara internal. Komentar yang sama dapat ditujukan untuk metode apapun, dari studi kasus sampai survai

acak. Tidak ada metode yang dapat menjamin validitas inferensi.

Shadish dkk (2002) berargumen bahwa validitas berkaitan erat dengan kebenaran sebab ada tiga prinsip filosofis yang mendasarinya, yaitu (a) *teori korespondensi* yang menyatakan bahwa sebuah klaim pengetahuan adalah benar jika klaim itu berhubungan dengan dunia kenyataan. Misalnya, klaim seseorang bahwa sekarang sedang hujan adalah benar kalau orang lain melihat keluar ruangan dan ternyata hujan sedang turun, (b) *teori koherensi* yang menyatakan bahwa sebuah klaim adalah benar jika klaim itu merupakan bagian dari sekumpulan klaim yang saling bertalian/koheren. Misalnya, klaim yang menyatakan bahwa mengisap mariyuana menyebabkan kanker adalah benar jika klaim itu konsisten dengan pengetahuan orang mengenai hasil mengisap mariyuana pada sistem hewan sama dengan pada sistem manusia, jika kanker merupakan akibat dari bentuk lain perilaku merokok, jika penyebab kanker meliputi sejumlah unsur yang diketahui mengikuti penghisapan mariyuana, serta jika mekanisme fisiologis yang menghubungkan penghisapan rokok dengan kanker juga diaktifkan ketika menghisap mariyuana, dan (c) *pragmatisme* yang menyatakan bahwa sebuah klaim adalah benar jika bermanfaat untuk meyakini klaim itu. Misalnya, kita mengatakan "elektron itu ada" jika penyimpulan wujud itu akan membawa makna atau kemampuan memprediksi sekumpulan pengamatan yang tanpanya lebih sulit dipahami. Elektron mungkin sesungguhnya tidak ada, namun dengan mempostulatkan bahwa elektron itu ada maka akan timbul ketertiban intelektual, serta ada manfaat praktis jika mengikuti kebiasaan yang memakai postulat itu. Shadish, dkk mengakui bahwa konsep validitas mereka dipengaruhi oleh

ketiga filosofi di atas. Validitas harus menyangkut korespondensi antara bukti empiris dengan inferensi teoretis. Validitas harus peka terhadap kadar sifat kesaling bertalian (koherensi) antara inferensi dengan teori dan temuan yang relevan. Validitas juga harus memiliki sifat pragmatis dalam menekankan kegunaan atas peniadaan penjelasan alternatif yang oleh peneliti diyakini dapat mengkompromikan klaim pengetahuan. Shadish, dkk. (2002) meyakini bahwa kebenaran dalam ilmu sosial memang merupakan konstruksi sosial namun juga tidak terlepas dari ketiga dasar filosofis tadi.

Tipologi validitas

Kategorisasi validitas eksperimen mengalami perkembangan sejak pertamakali ditulis dalam sebuah artikel oleh Campbell di tahun 1957 serta dalam buku karangan Campbell dan Stanley (1966). Campbell (1957) serta Campbell dan Stanley (1966) membagi validitas eksperimen menjadi dua jenis, yakni validitas internal (*internal validity*) dan validitas eksternal (*external validity*). Cook dan Campbell (1979) lebih memerinci pengertian validitas menjadi empat jenis: validitas konklusi statistik (*statistical conclusion validity*), validitas internal, validitas konstruk (*construct validity*) serta validitas eksternal. Shadish, Cook dan Campbell (2002) tetap mempertahankan keempat jenis validitas eksperimen tersebut seraya menitikberatkan hubungan keempatnya dengan upaya generalisasi pengambilan kesimpulan kausal.

Campbell (1957, hal. 297; Campbell & Stanley, 1966, hal. 5) menyatakan bahwa validitas internal berkaitan dengan pertanyaan "Apakah benar perlakuan eksperimental membuat perbedaan pada variabel eksperimen tertentu". Validitas internal

adalah *sine qua non*, yaitu dasar minimal yang harus dipenuhi oleh sebuah eksperimen sebab jika dasar minimal itu tidak dipenuhi maka sebuah eksperimen menjadi tidak bisa ditafsirkan. Validitas eksternal mempermasalahkan persoalan daya generalisasi: "Pada populasi, seting, variabel perlakuan, dan variabel pengukuran apa pengaruh tersebut dapat digeneralisasikan? (Campbell, 1957, hal. 297; Campbell & Stanley, 1966, hal 70)". Validitas eksternal terkadang disebut sebagai *generalizability* atau *representativeness*.

Cook dan Campbell (1979) mengembangkan lebih lanjut dua tipe validitas eksperimen tadi menjadi empat tipe validitas berdasarkan pada empat permasalahan praktis yang dihadapi peneliti, yaitu (a) Apakah ada hubungan antara variabel independen dan variabel dependen?, (b) Dengan adanya sebuah hubungan, apakah ada kemungkinan bersifat kausal dari sebuah variabel operasional ke variabel lain?, (c) Dengan kemungkinan adanya hubungan yang bersifat kausal dan diketahui dari satu variabel ke variabel lain, konstruk sebab-efek apakah yang terlibat dalam hubungan tersebut?, (d) Dengan kemungkinan adanya hubungan kausal dari konstruk A ke konstruk B, sejauhmana generalisasi hubungan tersebut pada semua orang, seting dan waktu?. Empat tipe validitas eksperimen tersebut adalah (a) validitas konklusi statistik. Validitas konklusi statistik mengacu pada kesesuaian penggunaan teknik statistik untuk membuat inferensi apakah variabel independen (perlakuan) dan variabel dependen (dampak perlakuan) berkorelasi, (b) validitas internal. Validitas internal mengacu pada apakah kovariansi variabel independen (perlakuan) dan variabel dependen (dampak perlakuan) merupakan akibat dari hubungan kausal, (c) validitas konstruk. Validitas konstruk adalah gene-

realisasi dari definisi operasional khusus ke konstruk yang lebih luas, serta (d) validitas eksternal. Validitas eksternal merupakan generalisasi sampel orang, seting, variabel independen serta variabel dependen yang digunakan oleh sebuah eksperimen ke populasi yang lebih luas berkaitan dengan permasalahan yang diteliti. Baik validitas konstruksi maupun validitas eksternal mengacu pada generalisasi.

Cook dan Campbell (1979) berpendapat bahwa kovariansi merupakan kondisi perlu untuk menentukan adanya hubungan kausal. Jika validitas internal mempertanyakan apakah terdapat hubungan kausal antara variabel independen dengan variabel dependen, maka berarti pula dapat dipertanyakan apakah terdapat kovariansi antara variabel independen dengan variabel dependen. Validitas konklusi statistik adalah pecahan dari validitas internal. Validitas konklusi statistik merupakan validitas mengenai kovariansi antara variabel independen dengan variabel dependen yang dihasilkan oleh bukti analisis statistik.

Pengertian validitas konstruk sebab-efek terkandung dalam kategori validitas eksternal-nya Campbell dan Stanley dibuku tahun 1966 serta dibuku-buku eksperimen ketika membahas "variabel pencemar (*confounding variable*)". Kajian mengenai variabel pencemar berkaitan dengan pertanyaan apakah efek (maksudnya perubahan dalam variabel dependen) itu disebabkan oleh variabel X (maksudnya manipulasi variabel independen) yang telah direncanakan atau apakah X dicemari oleh "harapan si eksperimenter" atau dikenal sebagai "efek Hawthorne" (Cook & Campbell, 1979). Validitas eksternal sudah dirumuskan sebagai generalisasi sampel orang, seting, variabel independen serta variabel dependen yang digunakan oleh sebuah eksperimen ke populasi yang

lebih luas dari sampel orang, seting, variabel independen serta variabel dependen. Validitas konstruk juga terkait dengan generalisasi sebab validitas konstruk mempertanyakan: "Dapatkah saya menggeneralisasikan dari satu rumusan operasional atau sejumlah rumusan operasional kepada sebuah konstruk acuan (*referent construct*)". Jadi validitas konstruk adalah pecahan dari validitas eksternal.

Shadish, dkk. (2002) tetap mempertahankan empat tipe validitas eksperimen di atas. Mereka merumuskan validitas konklusi statistik dan validitas internal sama dengan rumusan Cook dan Campbell (1979), namun mereka memodifikasi validitas konstruk dan validitas eksternal sesuai dengan umpan balik Cronbach bahwa dua jenis generalisasi kausal, yakni representasi dan ekstrapolasi, berlaku untuk semua unsur sebuah eksperimen, yaitu **unit** (*unit=u*), **perlakuan** (*treatment=t*), **observasi dampak perlakuan** (*observation=o*), dan **seting** (*setting=s*). **Unit** adalah unit terkecil atau biasanya dalam penelitian psikologi adalah individu, **perlakuan** adalah variabel independen yang dimanipulasi, **observasi dampak perlakuan** adalah pengukuran variabel dependen sebagai akibat dari perlakuan, serta **seting** yang menunjuk kepada lokasi/tempat/latar belakang tempat eksperimen. Keempat unsur yang ada dalam sebuah eksperimen tertentu disingkat **utos**. Generalisasi hubungan kausal yang pertama (representasi) adalah menggeneralisasikan dari unit, perlakuan, dampak perlakuan dan seting dimana data diperoleh dalam sebuah eksperimen (*utos*) ke konstruk lebih abstrak dan lebih tinggi yang direpresentasikan oleh *utos* tadi (konstruk sasaran, UTOS). Jadi jikalau validitas konstruk menurut Cook dan Campbell (1979) hanya menyangkut generalisasi dari konstruk perlakuan dalam

sebuah eksperimen (*t*) ke konstruk perlakuan yang lebih luas dan lebih tinggi (*T*) serta dari konstruk observasi dampak perlakuan yang terdapat dalam sebuah eksperimen (*o*) ke konstruk observasi dampak perlakuan yang lebih luas dan lebih tinggi (*O*), maka Shadish dkk (2002) memperluas cakupan validitas konstruk kesemua unsur yang ada dalam sebuah eksperimen (*utos*) yaitu dari konstruk *u* ke *U*, dari konstruk *t* ke *T*, dari konstruk *o* ke *O* serta dari konstruk *s* ke *S*.

Sebuah contoh mengenai validitas konstruk atau upaya membuat generalisasi dari *utos* ke *UTOS* digambarkan berikut: Apakah inferensi mengenai pengaruh kursus singkat tentang pembedahan dengan cara tur keliling rumah sakit yang dipimpin perawat (*t*) terhadap menurunnya pemakaian obat analgesik serta meningkatnya kegiatan hidup sehari-hari (*o*) para pasien bedah perut atau dada (*u*) di rumah sakit kota Montreal (*s*) dapat digeneralisasikan kepada konstruk yang lebih luas mengenai pendidikan kepada pasien (*T*) terhadap pemulihan fisik (*O*) para pasien bedah (*U*) di rumah sakit (*S*) (Shadish, dkk 2002 dan lihat Hastjarjo, 2010) Generalisasi dari *utos* ke *UTOS* dise-

but sebagai validitas konstruk.

Generalisasi kedua (ekstrapolasi) adalah menggeneralisasikan *utos* ke **UTOS* yang disebut sebagai validitas eksternal. *UTOS** (*UTOS* bintang) menunjukkan sampel subjek, perlakuan, pengukuran dampak perlakuan, dan seting yang tidak digunakan dalam sebuah eksperimen tertentu. Misalnya, apakah inferensi adanya pengaruh CBT versi Beck (*t*) terhadap menurunnya depresi yang diukur dengan BDI (*o*) pada mahasiswa UGM (*u*) yang mendatangi UKP Fakultas Psikologi UGM (*s*) dapat digeneralisasikan kepada CBT versi Ellis (*T**) terhadap menurunnya depresi yang diukur dengan Hamilton Depression Rating Scale (*O**) pada klien (*U**) di klinik psikologi Rumah Sakit X (*S**).

Ancaman terhadap validitas

Ancaman terhadap validitas adalah alasan khusus mengapa kita dapat salah ketika kita mengambil inferensi mengenai kovariansi, mengenai sebab-efek, mengenai konstruksi atau mengenai apakah hubungan kausal berlaku bagi variasi orang, seting, perlakuan dan dampak perlakuan.

Tabel 1.
Empat Tipe Validitas menurut Shadish, et.al (2002)

| |
|--|
| Validitas konklusi statistik: Validitas inferensi mengenai korelasi (kovariansi) antara perlakuan dengan dampak perlakuan. |
| Validitas internal: Validitas inferensi mengenai apakah kovariansi yang teramati antara perlakuan (A) dengan dampak perlakuan (B) mencerminkan sebuah hubungan kausal dari A ke B sebagaimana variabel tersebut dimanipulasi atau diukur. |
| Validitas konstruk: Validitas inferensi mengenai konstruk tingkat lebih tinggi yang merepresentasikan sampel khusus. |
| Validitas eksternal: Validitas inferensi mengenai apakah hubungan sebab-efek berlaku sepanjang variasi orang, seting, variabel perlakuan dan variabel pengukuran. |

Ancaman terhadap validitas eksperimen telah diidentifikasi melalui sebuah proses yang sebagian konseptual dan sebagian lagi empiris. Dasar empiris didapatkan dari komentar kritis kepada eksperimen yang telah dibuat sebelumnya, kebanyakan ancaman itu secara teoretis bersifat biasa. Ancaman yang berdasar empiris akan berubah seiring jaman oleh karena pengalaman akan kebutuhan ancaman baru dan keusangan ancaman lama. Misal, Shadish dkk (2002) menambah ancaman terhadap validitas konklusi statistik "estimasi besar efek yang tidak akurat" untuk mencerminkan realitas bahwa peneliti sosial sekarang menekankan estimasi besarnya efek kausal disamping uji signifikansi statistik. Sebaliknya, meskipun setiap ancaman terhadap validitas eksperimen memang terjadi dalam sebuah eksperimen, peluang terjadinya ancaman tersebut akan bervariasi dari satu konteks ke konteks lain. Daftar ancaman terhadap validitas merupakan alat bantu yang bersifat heuristik serta tidak secara universal berlaku dalam berbagai konteks penelitian.

Validitas konklusi statistik

Validitas konklusi statistik berkaitan dengan dua inferensi statistik yang saling berhubungan yang mempengaruhi komponen kovariansi inferensi kausal, yakni (a) apakah sebab dan efek berkovariansi. Dalam hal ini, orang dapat salah menyimpulkan bahwa sebab dengan efek berkovariansi padahal keduanya tidak berkovariansi (kesalahan tipe I) atau secara keliru menyimpulkan bahwa sebab dengan efek tidak berkovariansi padahal keduanya berkovariansi (kesalahan tipe II), serta (b) seberapa kuat keduanya berkovariansi. Kita dapat menaksir lebih tinggi atau lebih rendah besarnya kovariansi sebab-efek serta

taraf kepercayaan yang dimaksudkan oleh estimasi besarnya efek tersebut.

Cara yang banyak dipakai untuk menentukan apakah sebab dengan efek berkovariansi adalah melakukan pengujian signifikansi hipotesis nol (NHST=*null hypothesis significance testing*). Psikologi terbiasa memutuskan secara dikotomis: jika $p < 0,05$ maka ada perbedaan signifikan, sebaliknya jika $p > 0,05$ maka tidak ada perbedaan signifikan. Implikasi dari hasil analisis statistik yang tidak signifikan adalah kesimpulan bahwa sebab dengan efek tidak berkovariansi. Inferensi ini dapat salah dan mempunyai konsekuensi serius, oleh karenanya ancaman terhadap validitas konklusi statistik akan berkaitan mengenai mengapa seorang peneliti dapat keliru dalam membuat klaim menemukan tidak ada efek signifikan dengan menggunakan pengujian signifikansi hipotesis nol. Pengujian signifikansi hipotesis nol mempunyai permasalahan pro dan kontra (lihat Hastjarjo, 2000). Shadish dkk (2002) menulis perlunya menghindari penafsiran simplistik soal "ada efek" atau "tidak ada efek", dan percaya bahwa peran NHST akan semakin surut dalam ilmu sosial meskipun belum ada pengganti yang sempurna.

Ancaman terhadap Validitas Konklusi Statistik

Ancaman terhadap validitas konklusi statistik merupakan alasan mengapa peneliti mungkin salah dalam membuat inferensi mengenai adanya kovariansi antara dua variabel serta besarnya kovariansi antara dua variabel itu.

Berikut adalah daftar ancaman terhadap validitas konklusi statistik.

1. *Power statistik yang rendah*: Sebuah eksperimen dengan daya kekuatan

yang tak memadai kemungkinan akan keliru menyimpulkan bahwa hubungan antara perlakuan dengan dampak perlakuan tidak signifikan.

2. *Pelanggaran terhadap asumsi uji statistik*: Pelanggaran terhadap asumsi uji statistik dapat menghasilkan baik estimasi terlalu tinggi maupun estimasi terlalu rendah tentang besar dan signifikansi sebuah pengaruh.
3. *Permasalahan fishing and error rate*: Uji terhadap hubungan signifikan yang dilakukan berulang kali, jika tidak dikoreksi dapat membuat inflasi signifikansi statistik.
4. *Alat ukur yang tidak reliabel*: Kesalahan pengukuran akan melemahkan hubungan antara dua variabel dan memperkuat atau melemahkan hubungan antara tiga atau lebih variabel.
5. *Pembatasan kisaran*: Kisaran yang tereduksi pada sebuah variabel biasanya melemahkan hubungan antara variabel itu dengan variabel lain
6. *Ketidakreliabelan penerapan perlakuan*: Jika sebuah perlakuan yang diharapkan diterapkan dengan satu cara baku namun diterapkannya hanya sebagian untuk sejumlah responden, maka pengaruhnya mungkin lebih rendah dibanding dengan penerapan penuh.
7. *Varians luar seting eksperimen*: Sejumlah fitur satu seting eksperimen mungkin membuat inflasi kesalahan sehingga proses pendeteksian sebuah pengaruh akan lebih sulit.
8. *Heterogenitas unit*: Peningkatan variabilitas pada variabel dependen dalam kondisi perlakuan akan meningkatkan varians kesalahan sehingga proses pendeteksian sebuah pengaruh akan lebih sulit.
9. *Estimasi besarnya pengaruh yang tidak cermat*: Sejumlah teknik analisis statis-

tik secara sistematis membuat estimasi terlalu tinggi atau terlalu rendah terhadap besarnya sebuah pengaruh.

Validitas Internal

Validitas internal merujuk pada validitas inferensi mengenai apakah kovarians yang teramati antara A dan B mencerminkan sebuah hubungan kausal dari A ke B sebagaimana variabel tersebut dimanipulasi atau diukur. Validitas internal akan didukung jika peneliti dapat menunjukkan bahwa A mendahului B, bahwa A berkorelasi dengan B yang ditentukan oleh validitas konklusi statistik, serta tidak ada kemungkinan penjelasan lain yang dapat menerangkan hubungan antara A dan B.

Ancaman terhadap validitas internal

Berikut adalah daftar ancaman terhadap validitas internal, yaitu alasan-alasan mengapa inferensi bahwa ada hubungan kausal antara dua variabel mungkin tidak benar.

1. *Presedensi temporal yang kabur*: Kurangnya kejelasan variabel mana yang terjadi lebih dahulu mungkin akan menghasilkan kebingungan mengenai variabel mana yang menjadi penyebab dan mana yang menjadi efek/akibat.
2. *Seleksi*: Perbedaan sistematis ciri responden diantara kelompok eksperimen dan kontrol yang dapat juga menghasilkan efek yang teramati.
3. *Sejarah*: Kejadian yang berlangsung pada saat bersamaan dengan perlakuan dapat menghasilkan efek yang teramati.
4. *Maturasi*: Perubahan yang terjadi secara alamiah sepanjang waktu dapat keliru dikira sebagai efek perlakuan.

5. *Regresi*: Jika unit diseleksi berdasar skor ekstrim, mereka seringkali akan mempunyai skor yang kurang ekstrim pada variabel lain (termasuk skor pada pemberian tes ulang), sebuah kejadian yang dapat keliru dikira sebagai efek perlakuan.
6. *Atrisi atau mortalitas*: Mundurnya/hilangnya responden saat perlakuan atau saat pengukuran dapat menghasilkan efek artifaktual jika peristiwa kehilangan tersebut secara sistematis berkorelasi dengan kondisi perlakuan
7. *Pengujian*: Paparan sebuah tes dapat mengubah skor pada paparan tes selanjutnya, sebuah peristiwa yang dapat keliru dikira sebagai efek perlakuan.
8. *Instrumentasi*: Sifat-dasar sebuah instrumen pengukuran mungkin berubah dalam cara tertentu sesuai dengan perubahan waktu atau kondisi sehingga perubahan tersebut dapat keliru dikira sebagai efek perlakuan.
9. *Efek aditif dan interaktif ancaman terhadap validitas internal*: Pengaruh sebuah ancaman dapat ditambahkan dengan pengaruh ancaman lain.

Validitas konstruksi

Shadish dkk (2002, hal. 64) memberikan sebuah contoh mengenai persoalan validitas konstruk dalam penelitian psikologi yang diambil dari laporan *National Academy of Science* (Akademi Ilmu Pengetahuan Amerika):

“Dalam pengukuran bobot orang, hanya akan terjadi sedikit ketidaksepakatan mengenai arti dari konstruk yang diukur atau mengenai unit pengukuran (misal, sentimeter, gram, sentimeterkubik) Namun pengukuran pertumbuhan dalam ranah psikologis (misal, perbendaharaan

kata, penalaran kuantitatif, memori verbal, koordinasi mata-tangan, regulasi-diri) lebih bermasalah. Ketidaksepakatan mengenai definisi konstruk yang diukur akan lebih banyak muncul. Hal ini terjadi sebagian oleh karena seringkali tidak ada unit alamiah pengukuran (yaitu tidak ada yang dapat dibandingkan dengan penggunaan inci untuk mengukur tinggi)”.

Kutipan ini menunjukkan adanya dua permasalahan validitas konstruk dalam psikologi, yaitu pemahaman dan pengukuran konstruk.

Ancaman terhadap validitas konstruk

Ancaman terhadap validitas konstruk adalah alasan-alasan mengapa inferensi mengenai konstruk yang memberi ciri definisi operasional di sebuah penelitian mungkin keliru.

1. *Ketidaktepatan perumusan konstruk*: Kegagalan merumuskan sebuah konstruk mungkin menghasilkan inferensi yang keliru mengenai hubungan antara definisi operasional dengan konstruk.
2. *Pencemaran konstruk*: Definisi operasional biasanya mengandung lebih dari satu konstruk, dan kegagalan mendeskripsikan semua konstruk mungkin mengakibatkan inferensi mengenai konstruk yang tidak sempurna.
3. *Bias rumusan operasional-tunggal*: Definisi operasional sebuah konstruk baik yang kurang merepresentasikan konstruk yang bersangkutan maupun yang mengukur konstruk yang tidak relevan akan menyebabkan kerumitan inferensi.
4. *Bias-metode tunggal*: Jika semua definisi operasional menggunakan metode yang sama (misal, laporan-diri) maka

metode itu sendiri merupakan bagian dari konstruk yang dipelajari.

5. Pencemaran konstruk dengan level konstruk: Inferensi mengenai konstruk yang paling tepat merepresentasikan definisi operasional penelitian mungkin gagal mendeskripsikan level tertentu konstruk yang diteliti secara aktual.
6. Struktur faktorial yang peka terhadap perlakuan: Struktur sebuah pengukuran mungkin berubah sebagai akibat satu perlakuan, perubahan tersebut ada kemungkinan tersembunyi jika metode penyekoran yang sama selalu digunakan.
7. Perubahan laporan-diri yang bersifat reaktif: Laporan-diri dapat dipengaruhi oleh motivasi partisipan untuk berada didalam kondisi perlakuan, namun motivasi mungkin berubah sesudah penempatan subjek dilakukan.
8. Reaktivitas terhadap situasi eksperimen: Respons partisipan mencerminkan bukan hanya perlakuan dan pengukuran melainkan juga persepsi partisipan akan situasi eksperimen, serta persepsi itu merupakan bagian dari konstruk perlakuan yang sebenarnya diteliti.
9. Harapan eksperimenter: Eksperimenter dapat mempengaruhi respons partisipan dengan mengirimkan harapan akan respons yang diinginkan, serta harapan itu merupakan bagian konstruk perlakuan yang sebenarnya diteliti.
10. Efek kebaruan dan gangguan: Partisipan mungkin merespon secara luar-biasa baik terhadap sebuah inovasi yang baru atau secara luar biasa jelek terhadap situasi yang mengganggu rutinitas mereka, sebuah respons yang kemudian harus dimasukkan kedalam bagian deskripsi konstruk perlakuan.
11. Penyamaan kompensatoris: Jika perlakuan memberikan barang atau jasa yang menyenangkan, administrator, staf, atau konstituen mungkin menyediakan barang atau jasa kompensasi terhadap mereka yang tidak mendapatkan perlakuan, maka tindakan seperti ini kemudian wajib dijadikan bagian dari deskripsi konstruk perlakuan.
12. Persaingan kompensatoris: Partisipan yang tidak mendapatkan perlakuan mungkin termotivasi untuk menunjukkan bahwa mereka mampu berperilaku sebaik partisipan yang memperoleh perlakuan, dan tindakan ini kemudian wajib dijadikan bagian dari deskripsi konstruk perlakuan.
13. Demoralisasi kebencian: Partisipan yang tidak mendapatkan perlakuan mungkin begitu benci atau mengalami demoralisasi sehingga mereka mungkin merespon secara lebih negatif dibandingkan seperti biasanya, dan tindakan demoralisasi kebencian ini kemudian wajib dijadikan bagian dari deskripsi konstruk perlakuan.
14. Difusi perlakuan: Partisipan mungkin mendapatkan jasa dari kondisi perlakuan yang bukan tempatnya mereka ditempatkan, hal ini membuat deskripsi konstruk dari kedua kondisi perlakuan sulit digambarkan.

Validitas Eksternal

Validitas eksternal merujuk pada inferensi sejauhmana sebuah hubungan kausal berlaku sepanjang variasi orang, seting, perlakuan serta dampak perlakuan (Shadish dkk, 2002). Validitas eksternal mencakup permasalahan hubungan kausal berlaku (a) sepanjang variasi orang,

seting, perlakuan dan dampak perlakuan yang terdapat dalam sebuah eksperimen yang telah dilaksanakan serta (b) bagi orang, seting, perlakuan dan dampak perlakuan yang tidak terdapat dalam sebuah eksperimen yang telah dilaksanakan. Sasaran generalisasi dapat bermacam-macam, misalnya (a) dari sempit ke luas. Misalnya, sebuah eksperimen mengenai penataan waktu kerja yang dilakukan di propinsi DIY apakah dapat diberlakukan secara umum ke populasi negara Indonesia?, (b) Dari luas ke sempit. Dari sampel yang dipakai sebuah eksperimen apakah dapat diberlakukan kepada kelompok yang lebih kecil atau malah ke pribadi seseorang? Misalnya, terapi kognitif keperilakuan yang mampu menurunkan kecemasan penyintas gempa bumi Aceh, apakah dapat diterapkan kepada seseorang penyintas gempabumi yang berdomisili di kota Banda Aceh, (c) Pada level sama. Dari sampel yang dipakai sebuah eksperimen ke sampel lain yang level agregasinya setingkat. Misal, sebuah eksperimen untuk meningkatkan komitmen kerja yang dilakukan di sebuah perusahaan swasta diberlakukan pada sebuah perusahaan swasta lain dengan tipe dan besar organisasi sama, (d) Kepada jenis yang sama atau berbeda. Sasaran generalisasi dari kasus a, b dan c di atas mungkin sama dengan sampel sebuah eksperimen (misalnya hasil eksperimen dengan karyawan laki-laki di perusahaan tekstil X ke karyawan laki-laki di perusahaan tekstil Y) atau berbeda dengan sampel sebuah eksperimen (misalnya hasil eksperimen dengan orang Jawa di kota Yogyakarta ke orang Sunda di kota Bandung), (e) Sampel acak ke anggota populasi. Dalam kasus yang jarang terjadi, generalisasi dibuat dari sampel hasil pemilihan secara acak ke anggota populasi darimana sampel tersebut diacak.

Ancaman terhadap validitas eksternal

Estimasi sejauhmana sebuah hubungan kausal berlaku sepanjang variasi orang, seting, perlakuan dan dampak perlakuan secara konseptual mirip dengan pengujian interaksi statistik. Misalnya jika ada interaksi antara perlakuan pendidikan dengan kelas sosial anak, maka peneliti tidak dapat mengambil kesimpulan bahwa hasil yang sama akan berlaku pada semua kelas sosial anak. Oleh sebab itu ancaman terhadap validitas eksternal dikemas dalam interaksi hubungan kausal dengan unit, perlakuan, dampak perlakuan dan seting. Konsep yang melatarbelakangi interaksi itu lebih penting: mencari cara mengapa sebuah hubungan kausal dapat atau tidak dapat berubah sepanjang variasi orang, perlakuan, dampak perlakuan dan seting.

Ancaman terhadap validitas eksternal adalah alasan-alasan mengapa inferensi mengenai hasil eksperimen dapat berlaku sepanjang variasi orang, seting, perlakuan dan dampak perlakuan mungkin salah. Berikut adalah daftar ancaman terhadap validitas eksternal sebuah eksperimen.

1. Interaksi antara hubungan kausal dengan unit: sebuah efek yang ditemukan dengan jenis unit tertentu mungkin tidak berlaku seandainya unit lain yang telah diteliti.
2. Interaksi antara hubungan kausal dengan variasi perlakuan: sebuah efek yang ditemukan dengan sebuah variasi perlakuan tertentu mungkin tidak berlaku dengan variasi lain perlakuan tersebut, atau jika perlakuan itu dikombinasikan dengan perlakuan lain, atau jika hanya sebagian dari perlakuan itu digunakan.
3. Interaksi antara hubungan kausal dengan dampak perlakuan: Sebuah efek yang ditemukan dengan satu jenis

observasi dampak perlakuan mungkin tidak berlaku jika jenis lain observasi dampak perlakuan digunakan.

4. Interaksi antara hubungan kausal dengan seting: Sebuah efek yang ditemukan dengan satu jenis seting mungkin tidak berlaku jika jenis seting lain digunakan.
5. Mediasi tergantung konteks: sebuah mediator yang menjelaskan hubungan kausal dalam satu konteks mungkin tidak akan memediasi dalam konteks lain.

Penutup

Konsep validitas eksperimen berbeda dengan konsep validitas pengukuran. Validitas eksperimen menggambarkan ancar-ancar kebenaran inferensi yang dibuat peneliti dalam satu eksperimen. Tipologi validitas sebuah eksperimen terdiri dari empat (a) validitas konklusi statistik, (b) validitas intenal, (c) validitas konstruk, serta (d) validitas eksternal. Inferensi yang dibuat peneliti mengenai sebuah eksperimen bisa salah. Alasan-alasan yang membuat inferensi mengenai sebuah eksperimen salah disebut sebagai ancaman terhadap validitas.

Daftar Pustaka

- Affandi, G. R., & Hastjarjo, T. D. (2010). Pengaruh tipe penentuan tujuan (*Goal-Setting*) terhadap performansi akademik bahasa Inggris siswa dengan efikasi diri dan kemampuan awal bahasa Inggris sebagai kovariabel. *Jurnal Psikologi Tabularasa*, 5(2), 264-368.
- Azwar, S. (1997). *Reliabilitas dan validitas*. Edisi ketiga. Yogyakarta: Pustaka Pelajar.
- Campbell, D. T. (1957). Factor relevant to the validity of experiments in social settings. *Psychological Bulletin*, 34(4), 297-312.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Co.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin Co.
- Hadi, S. (1977). *Metodologi research: Untuk penulisan paper, skripsi, tesis dan disertasi*. Jilid dua. Yogyakarta: Yayasan Penerbitan Fakultas Psikologi UGM.
- Hastjarjo, T. D. (2000). Seputar kontroversi uji signifikansi hipotesis nihil. *Buletin Psikologi*, 8(2), 1-6.
- Hastjarjo, T.D. (2011). Kausalitas menurut tradisi Donald Campbell. *Buletin Psikologi*, 19(1), 1-43.
- Kerlinger, F.N. (1986). *Foundations of behavioral besearch*. 3rd Edition. New York: Holt, Rinehart and Winston.
- Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral besearch*. 4th Edition. Fort Worth: Harcourt College Publisher.
- Masrun. (1975). *Pengukuran dalam pendidikan*. Yogyakarta: Gadjah Mada University Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Co.
- Suryabrata, S. (2000). *Pengembangan alat ukur psikologis*. Yogyakarta: Penerbit Andi.