



VALIDITAS, RELIABILITAS, GENERALITAS DALAM RISET KEBAHASAAN

MK. QUANTITATIVE RESEARCH METHODOLOGY

ROHMANI NUR INDAH

PENGERTIAN VALIDITAS

Validitas adalah pertimbangan yang paling penting dalam mengembangkan dan mengevaluasi pengukuran instrumen. Apakah instrumen cukup **handal** untuk digunakan dalam mengumpulkan data? Inilah yang menjadi fokus validitas . Artinya interpretasi dan makna dari skor yang didapat dari suatu instrumen apakah sudah sesuai dengan tujuan dibuatnya instrumen tersebut. Standar terbaru untuk American Psychological Association, mendefinisikan validitas sebagai "sejauh mana" bukti dan teori mendukung interpretasi skor yang diperoleh dari suatu instrumen tes.

Misalnya, tes kosakata Bahasa Inggris kelas tujuh yang memungkinkan seorang guru untuk membuat penafsiran valid tentang pencapaian kosakata Bahasa Inggris siswa kelas tujuh, tapi tidak menghasilkan interpretasi yang valid tentang kemampuan siswa kelas tujuh untuk menulis paragraf berbahasa Inggris. Adapun jika guru tersebut mencoba menggunakan tes menjodohkan kalimat acak dalam Bahasa Inggris menjadi instrumen untuk mengukur kemampuan menulis siswa, hal ini juga

termasuk tidak valid. Sehingga skor yang diperoleh siswa dalam tes menjodohkan kalimat tersebut tidak bisa digunakan untuk menafsirkan kemampuan menulis siswa.

Dalam hal ini yang perlu diingat yaitu sebuah tes mungkin valid untuk digunakan dengan satu populasi tetapi tidak dengan yang lain. Karena itu dibutuhkan uji validitas yang memadai sebelum tes digunakan sebagai instrumen dalam pengumpulan data. Uji validitas sangat beragam, Cohen et al. (2007) menyebutkan antara lain:

- content validity
- criterion-related validity
- construct validity
- internal validity
- external validity
- concurrent validity
- face validity
- jury validity
- predictive validity
- consequential validity
- systemic validity
- catalytic validity
- ecological validity
- cultural validity
- descriptive validity
- interpretive validity
- theoretical validity
- evaluative validity

Apakah perlu diukur semuanya? Tentunya akan memberatkan peneliti, meskipun lebih banyak aspek yang digunakan untuk menguji validitas instrumen akan lebih baik, namun bisa dipilih sesuai tujuan penelitian. Utamanya yang diuji adalah validitas berdasarkan (a) konten, (b) kriteria, dan (c) konstruk instrumen.

- (a) Pada validitas berdasarkan konten yang diuji berkaitan dengan apakah butir-butir pernyataan (item-item) yang tersusun dalam kuesioner atau tes sudah mencakup semua materi yang hendak diukur. Jika penelitian akan melihat tentang kompetensi pragmatik calon guru Bahasa Inggris maka diperlukan kajian literatur (literature review) kompetensi pragmatik penutur Bahasa Inggris. Berdasarkan literature review, baru kemudian menyusun kuesioner dalam

beberapa bagian: Bagian 1, Informasi latar belakang responden; Bagian 2, pengetahuan sosiolinguistik; Bagian 3, pengetahuan kaidah bahasa; Bagian 4, kemampuan memahami konteks tutur. Pernyataan-pernyataan dalam kuesioner disusun berdasarkan masing-masing kompetensi pragmatik agar setiap item dapat mewakili seluruh landasan teoretis tentang topik penelitian tersebut. Dari segi konten, juga dikenal face validity atau validitas tampak muka, yang mengukur sejauh mana instrumen mengukur apa yang semestinya diukur. Dalam hal ini, dari segi tampilan yang ditanyakan apakah tes tampaknya handal, misalnya dalam tes tata bahasa tidak akan berisi soal tes tentang sejarah bahasa.

(b) Berdasarkan kriteria, validitas menguji sejauh mana skor yang diperoleh memiliki relasi secara sistematis pada kriteria hasil yang ditetapkan. Instrumen tes digunakan untuk menyimpulkan sebegus apa performa berdasarkan kriteria tersebut. Kriteria mutlak bersifat relevan, reliabel atau dapat dipercaya, serta bebas bias. Validitas kriteria berdasarkan waktunya terdiri dari (1) concurrent (bersamaan) dan (2) prediktif.

(1) Yang dimaksud validitas konkuren yaitu hubungan asesmen dan skor yang diperoleh pada waktu yang sama. Misalnya peneliti telah mengembangkan tes bakat bahasa asing dan membutuhkan bukti bahwa tes tersebut benar-benar mengukur kemampuan bahasa asing. Peneliti bisa memilih memakai tes bakat bahasa asing yang terkenal dan telah divalidasi sebelumnya, lalu menerapkan tes itu kepada sekelompok siswa, dengan menentukan korelasi antara dua set skor. Korelasi substansial antara yang tes baru dan tes umum dapat menjadi bukti bahwa tes baru tersebut juga mengukur bakat bahasa asing. Kriteria lain yang tersedia pada saat itu mungkin saat ini.

(2) Validitas prediktif memastikan hubungan antara skor pada asesmen dan skor kriteria mendatang. Misalnya pada tes bakat bahasa asing, orang akan melihat pada hubungan antara skor pada tes bakat skolastik untuk memprediksi performa bahasa asing siswa.

(c) Selain berdasarkan konten dan kriteria, validitas juga diuji berbasis konstruknya atau diukur berdasarkan kerangka (construct) teoretis yang tepat dan relevan. Kuesioner yang memiliki validitas konstruk tinggi selalu berdasarkan definisi atau batasan para ahli tentang konsep tersebut, bukan pada definisi kamus. Misalnya, tujuan penelitian mengukur hubungan kompetensi pragmatik guru Bahasa Inggris terhadap kemampuan menyimak siswa kelas delapan, maka perlu ditentukan dulu konsep teoretis tentang teori kompetensi pragmatik serta hubungannya dengan kemampuan menyimak. Berdasarkan batasan-batasan tersebut kemudian disusun butir-butir pernyataan yang sesuai. Jika pada sebagian isi instrument tidak mencerminkan tentang kedua konsep tersebut, maka validitas konstruknya tidak ditemukan.

Selain konstruk teoretis, juga dikenal konstruk psikologis, yaitu pada validitas construct-related evidence. Dalam hal ini yang diuji yaitu sejauh mana hasil skor tes merefleksikan konstruk psikologis yang diukur. Apa saja yang termasuk konstruk psikologis? Di antaranya yaitu intelegensi, motivasi, kecemasan, atau berpikir kritis. Konstruk pertanyaan dalam instrumen semestinya menyertakan aspek ini secara lebih komprehensif.

Pada intinya validitas harus setia pada premis penelitian yang dalam hal ini untuk mendukung prinsip-prinsip berikut:

- dapat dikontrol
- dapat direplikasi
- dapat diprediksi
- disimpulkan menjadi pernyataan universal mengenai suatu perilaku
- bebas konteks
- merupakan fragmentasi penelitian
- pengacakan sampel

- teramati

Mengapa prinsip tersebut mengemuka? Karena penelitian kuantitatif **berbeda** dari penelitian kualitatif yang prinsipnya sebagai berikut:

- tidak dikendalikan karena memiliki setting alami atau data natural
- tidak bebas konteks, melainkan terikat pada konteks tertentu
- data tidak acak karena data dipilih berdasarkan situasi sosial dan budaya
- simpulan tidak serta merta bersifat universal karena tergantung sudut pandang peneliti sebagai instrumen utama penelitian
- tidak dapat diprediksi secara umum karena sifat analisisnya penuh penafsiran hermenetik
- tidak bersifat fragmentasi melainkan memandang penelitian sebagai holisme yang melibatkan berbagai sistem dan sudut pandang sebagai kesatuan yang utuh
- penelitian pada data deskriptif, bukan skala atau variabel terukur
- memperhatikan proses tidak hanya hasil penilaian
- data dianalisis secara induktif, tidak hanya pada kategori tertentu
- data tergantung pada responden penelitian, tidak dibatasi peneliti
- memahami fenomena dengan melibatkan sudut pandang partisipan, tidak hanya peneliti
- validasi responden dipandang penting
- tujuan utama untuk menangkap makna dan memahami maksud

PENGERTIAN RELIABILITAS

Pada dasarnya reliabilitas mengukur seberapa jauh instrumen dapat dipercaya karena sifatnya yang **ajeg** atau konsisten. Seperti timbangan badan, jika digunakan seseorang selama beberapa kali menunjukkan skala yang berbeda, artinya hasil pengukuran berat badan tidak dapat dipercaya, atau timbangan tersebut tidak reliabel. Jika hasil tes diskor oleh dua orang penilai hasilnya berbeda jauh, perlu diragukan tingkat reliabilitasnya. Jika suatu tes kosakata diskoring pada hari senin dengan hari selasa, standar pemberian skor ajeg maka disebut reliabel, namun jika berbeda maka artinya kurang reliabilitasnya.

Dalam penelitian kuantitatif, terdapat dua cara umum yang digunakan banyak peneliti untuk menentukan tingkatan reliabilitas: (1) Test-retest reliability atau pengukuran ulang dan (2) tes konsistensi internal (Internal consistency). Pada tes ulang, seorang peneliti melakukan tes pada sampel penelitian yang sama dalam waktu yang berbeda. Dengan kata lain, seorang responden diberi kuesioner dan/atau tes yang sama dalam waktu yang berbeda, lalu hasil dari kedua tes tersebut dikorelasikan (Pearson Product moment) untuk melihat konsistensi hasil. Semakin tinggi tingkat korelasi pada tes pertama dan kedua, makin baik reliabilitas skala pengukuran. Namun demikian cara ini dipandang tidak efisien. Lebih disarankan menggunakan tes konsistensi internal yang ditunjukkan dengan koefisien Alpha (Cronbach Alpha) dengan rentang 0 (tidak memiliki reliabilitas) sampai 1 (reliabilitas sempurna). Selain itu juga bisa menggunakan Spearman-Brown rank order yang dilakukan dengan melakukan uji belah dua, atau membagi instrument tes kedalam dua bagian untuk dikorelasikan. Koefisien korelasi sempurna yaitu 1.00

Ada tiga prinsip yang harus dicermati dalam reliabilitas yaitu stabilitas, ekuivalen, dan konsistensi internal. Prinsip stabilitas artinya penilaian dari instrumen hasilnya stabil lintas waktu pada

responden yang sama. Artinya skoring pada satu sampel meskipun diskor hari ini atau pecan depan, hasilnya tetap sama. Mengenai prinsip ekuivalen, artinya penilaian konsisten dalam mengukur suatu variabel dalam kriteria dan tingkat kesulitan yang setara. Misalnya pre-tes dan pos-tes memiliki konsistensi skoring item secara ekuivalen. Pada saat skoring juga bisa melibatkan inter-rater, yaitu dinilai lebih dari satu orang peneliti untuk melihat ekuivalen tidaknya. Jika perbedaan skor antara dua rater terlalu lebar, bisa mengundang rater atau penilai ke tiga. Pada prinsip konsistensi internal artinya pada satu instrumen tes memiliki proporsi tingkat kesulitan yang konsisten. Teknik uji konsistensi internal sebagaimana yang dijelaskan pada paragraf sebelumnya.

Generalitas

Apa itu generalitas atau generalisasi dalam penelitian kuantitatif? Generalisasi adalah suatu kenyataan kebenaran yang terjadi dalam suatu realitas tentang suatu masalah yang diperkirakan akan berlaku pada suatu populasi tertentu. Generalisasi dapat dihasilkan melalui suatu metode perkiraan / metode estimasi yang umum berlaku di dalam statistika induktif.

Jika dalam konteks penelitian kebahasaan, generalisasi dilakukan setelah mendapatkan kenyataan kebenaran yang terjadi dalam suatu realitas pada sampel, sehingga fenomena yang sama diperkirakan akan berlaku juga pada populasi tertentu yang diwakili oleh sampel yang ditentukan. Misalnya pada satu sampel mahasiswa jurusan Bahasa Inggris universitas A yang telah mengikuti pelatihan dengan intruksi eksplisit pemodelan berpikir kritis menunjukkan peningkatan yang signifikan pada nilai mata kuliah critical reading dan critical writing. Maka jika pada populasi mahasiswa jurusan Bahasa Inggris di seluruh universitas pada kota A tersebut, diprediksi bahwa pemberian pelatihan dengan intruksi eksplisit pemodelan berpikir kritis juga akan memberikan peningkatan yang signifikan pada nilai critical reading dan critical writing.

Membincang generalitas dalam penelitian kuantitatif kebahasaan erat kaitannya hasil uji validitas dan reliabilitas. Jika instrumen yang digunakan tidak valid dan reliabel, maka hasil analisis dengan data yang diperoleh dari instrument tersebut sangat lemah untuk digeneralisir pada populasi yang lebih luas. Selain itu perlu juga diketahui beberapa pengukuran varians (measurement of variability) untuk mengestimasi mean dan variansi skor suatu tes. Prinsip dasar dari model generalisasi validitas yaitu mengenali apakah muncul error sistematis yang berhubungan dengan kekurangan pada proses statistik dan error dari random sampling.

Dengan melakukan uji generalisasi validitas, manfaatnya yaitu untuk mengetahui sejauh mana suatu instrument dapat berfungsi untuk menarik kesimpulan ilmiah tentang hubungan antara variabel. Misalnya jika seorang guru Bahasa Inggris membuat perangkat tes untuk menguji kompetensi pragmatic siswa kelas sembilan, maka hasil uji generalitas validitas tes itu jika baik artinya tes pragmatic tersebut handal digunakan untuk siswa kelas sembilan tidak hanya pada sampel penelitian tapi juga pada siswa dalam cakupan populasi yang lebih luas.

Sumber:

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education*. Cengage Learning.

Cohen, L., Manion, L., & Marrison, K. (2007). *Research in education sixth edition*.

Creswell, J. W. (2012). *Educational research: planning, Conducting, and Evaluating*.