

**VALIDITAS SEBAGAI ALAT PENENTUAN
KEHANDALAN TES HASIL BELAJAR**

Siswanto¹

Abstrak

Penentuan keberhasilan pembelajaran sangat penting untuk dapat mengetahui tingkat ketercapaian tujuan pembelajaran. Untuk mengetahui tingkat ketercapaian dari tujuan pembelajaran sangat diperlukan alat yang dapat memberikan gambaran ketercapaian pembelajaran serta keterserapam materi yang diberikan. Untuk itu ulah diperlukan alat tes untuk mengetahui hal tersebut, dengan adanya alat tes belum dapat menjamin kita akan mendapatkan gambaran ketercapaian pembelajaran apabila alat yang kita gunakan tidak dapat diandalkan. Oleh sebab itu reliabilitas menjadi hal yang penting untuk disyaratkan bagi pembuat alat tes.

Beberapa alternatif validitas yang dibahas dalam tulisan ini adalah validitas isi, validitas kriteria, validitas konstruk. Dalam validitas isi diuraikan alasan serta berbagai kriteria yang harus dipenuhi dalam validitas isi dan pemaparan yang terakhir adalah validitas tampak akan dikupas sedikit terkait dengan kemungkinan penggunaan jenis validitas ini. Validitas kriteria membahas bagian dari validitas kriteria yaitu validitas prediktif dan validitas konkrue serta berbagai faktor yang mempengaruhinya. Dalam validitas konstruk dipaparkan tujuan dan alasan penggunaan validitas konstruk serta langkah-langkah penentuan validitas konstruk.

Dari beberapa jenis validitas tersebut dapat dijadikan alternatif bagi pembuat alat tes untuk menentukan kehandalan alat yang digunakan, sehingga hasil dari alat tes tersebut benar-benar memberikan gambaran yang terkait dengan ketercapaian keberhasilan pembelajaran.

Kata Kunci : Alat tes, Validitas Tes, Kehandalan alat tes

PENDAHULUAN

Pengajar sering kali memberi instrument yang terdiri dari kegiatan yang diluar konteks dengan yang di contohkan, serta berusaha membuat beberapa pendapat tentang aktivitas murid yang berada di luar materi ujian. Bagaimanapun itu tidak baik jika murid melakukan ujian dengan beberapa kompetensi sekaligus, bagaimana mereka akan

¹ Staf Pengajar Jurusan Pendidikan Akuntansi – Universitas Negeri Yogyakarta

memiliki kompetensi. Di lain tempat ada yang memiliki anggapan, siapa yang memiliki nilai kecerdasan individu yang tinggi diharapkan akan memiliki pengaruh yang baik pada jabatan/pekerjaan serta kecakapan dalam pendidikan. Harapan dalam study sosial memiliki banyak harapan untuk berbagai pengetahuan, kemampuan serta kesamaan dalam berbagai langkah yang pada dasarnya memiliki sifat-sifat yang sama antara kecerdasan dan kemampuan membaca.

Instruksi yang efektif tergantung dari bagian kemampuan pengajar mendeteksi pengetahuan yang sulit dan dengan pilihan cara yang efektif dari berbagai cara yang ada, sehingga tepat dan pantas untuk memperbaiki ketercapaian materi yang disajikan. pengajar memerlukan hal yang dapat dipercaya, relevan dan memiliki banyak referensi yang terkait dengan keberhasilan siswanya. Informasi yang benar dapat membantu seorang pengajar dalam menentukan tindakan dalam setiap pertemuan dalam pembelajaran, perbedaan dan peningkatan yang sesuai dengan kurikulum akan lebih baik apabila didasarkan atas dasar kenyataan yang terjadi dari pada subjektifitas pengajar.

Kebenaran dalam menentukan langkah yang memiliki keluasaan dapat digunakan dalam membuat suatu keputusan, Sebagai contoh akan dapat meningkatkan hasil yang maksimal jika pengajar dapat memperkirakan jumlah siswa yang akan berhasil dalam proses pembelajaran.

Dari tingkatannya langkah yang telah ditentukan dapat dimanfaatkan secara efektif, sehingga jangan sampai keputusan yang dibuat dan memberikan pemahaman yang salah dan tidak relevan, mereka tidak memiliki ketepatan. Bagaimana tidak suatu hal dapat dipercaya oleh mereka, jika langkah yang dapat digunakan tidak benar untuk mencapai tujuan yang diinginkan.

Kebenaran tidak bisa dibuat dengan mengarang akan tetapi memerlukan berbagai bukti-bukti yang mendukung serta standart untuk pelaksanaan ujian tersebut. Untuk mendukung dengan bukti pelaksanaan ujian yang telah diberikan serta sertifikasi kelulusan untuk mereka yang berupa gelar.

Sebagai contoh ujian ini terdiri dari 75 perkalian, dan 25 pokok yang membahas perbandingan bilangan, menurut perkiraan langkah yang perlu diambil untuk mendapatkan pengetahuan yang diinginkan harus disesuaikan dengan kenyataan hasil serta referensi yang digunakan serta kondisi tingkat pendidikan siswa.

Meskipun banyak tuntutan adanya referensi yang dapat diaju, Borman dan Sander memberikan referensi yang baik untuk dapat memaknai tes serta langkah – langkah praktis yang dapat menentukan tingkat kepercayaan yang tinggi. Pemberian informasi tentang teknik untuk menghadapi tes mengarang perlu dihindari, antara lain distribusi normal populasi. Cara ini digunakan dalam perhitungan bukti-bukti dari substansi yang dapat dipercaya untuk digunakan dalam memenuhi ketepatan yang diinginkan, tetapi harus disesuaikan dengan standar tes yang telah ditentukan oleh penulisnya.

Saat adanya ukuran yang tidak terbaca saat adanya ujian, ini kanan menyebabkan adanya fakta yang tidak terungkap dari pokok bahasan yang diujikan. Dalam semua kebenaran, seharusnya dicatat berbagai pokok bahasan yang diujikan, dimana tidak satupun pokok bahasan yang mengharuskan pengetahuan murid yang bebas dari tingkat penerikan kembali atau akan munculnya ilmu yang modern dari hasil pembahasan pokok bahasan tersebut, yang dapat digunakan untuk proses pembelajaran yang tepat. Jika Borman dan Sander mengemukakan salah satunya adalah ringkasan pertemuan dengan instruktur dari langkah yang telah disepakati, didalamnya mereka itu akan menemukan

beberapa contoh setiap kesalahan dalam tes serta perbaikan dan konstruksi yang dijalankan dan mengatur sesuai dengan kaidah yang berlaku.

METODE EMPIRIK MENENTUKAN VALIDITAS VALIDITAS ISI (*CONTENT*)

Keputusan berdasarkan *validitas isi* menentukan apakah siswa telah menguasai, terkelompok dalam, atau butir yang gagal mengukur tujuan pembelajaran khusus. Butir soal memiliki validitas isi jika mengarahkan siswa untuk mendemonstrasikan keterampilan dan kompetensi yang dibutuhkan sesuai dengan tujuan pembelajaran. Tes memiliki validitas isi jika perilaku dan subyek matter sesuai dengan butir yang berhubungan dengan tujuan khusus. Masuknya butir soal yang tidak sesuai dengan tujuan atau gagal memasukan butir soal yang sesuai dengan tujuan akan menghasilkan isi yang *tidak valid*.

Beberapa tes diarahkan mengukur tujuan secara langsung, walaupun keterampilan atau kemampuan yang diukur secara terwujud. Misalnya mengukur kemampuan mengeja dengan *menyeleksi* kata-kata yang salah eja dibandingkan dengan yang dieja dengan benar. Kemampuan untuk mengidentifikasi kata-kata yang salah eja berbeda dengan kemampuan mengeja kata secara lisan dengan benar. Contoh lain adalah tes prestasi membaca lambang-lambang matematika di California dimana siswa diminta untuk mendefinisikan kata "*star*" dan "*choose*" yang relevan dengan bahasa matematika seperti *pembagi*, *mengurangi*, dan *hasil bagi*. Kemampuan untuk mendefinisikan bahasa matematika harus disesuaikan dengan istilah yang lebih jelas untuk mengukur tujuan pembelajaran.

Kriteria yang dapat diterima untuk validitas isi membutuhkan 2 kelompok independent atau individu dalam mengkonstruksi tes menggunakan spesifikasi yang sama. Setiap hal harus diberikan informasi seperti (1) deskripsi umum dari butir yang dikonstruksi; (2) rangkaian tujuan perilaku; (3) jumlah butir yang dikonstruksi untuk setiap tujuan pembelajaran; (4) format (*multiple choice*, melengkapi, esay dan lain-lain); (5) sifat siswa yang mengikuti tes harus dipersiapkan (umur, kemampuan membaca, latar belakang sosial ekonomi, dan data relevan lain); dan (6) tujuan terhadap pengadministrasian dan penskoran tes. Ketika 2 tes telah dikonstruksi, kedua bentuk diberikan kepada kelompok siswa yang sama; korelasi antara 2 pengukuran akan menunjukkan seberapa baik butir soal mengukur kemampuan atau sifat yang sama. Namun demikian Lee Cronbach (1971) menyatakan bahwa kesulitan yang melekat pada validitas tes melalui prosedur ini telah mendorong kebanyakan individu untuk melakukan usaha dan hanya sedikit usaha tersebut yang tidak memberikan kepuasan.

Biasanya pengajar akan mengevaluasi validitas isi dari tesnya sendiri dan hal ini dimulai pada tahap perencanaan. Karena pengajar terlibat langsung dengan tes yang dibuat maka ada manfaat dengan bertanya pada pengajar lain untuk membandingkan butir soal yang sesuai dengan tujuan untuk menghindari soal-soal yang bias. Dalam mempertahankan masuknya suatu butir dalam tes atau menjelaskan hilangnya butir yang tidak sesuai tujuan akan mendorong pengajar lebih mengkonstruksi bentuk tes yang lebih bermakna. Proses mengevaluasi validitas isi dari tes menghasilkan hubungan yang sangat erat antara mengajar dan menguji.

Mengembangkan validitas dengan menentukan hubungan antara butir soal dan tujuan utamanya diterapkan dalam *tes prestasi* atau mengukur perolehan siswa dari

subyek matter. Umumnya pengajar membuat tes tidak mengukur berdasarkan kriteria tetapi berdasarkan tujuan pengajar itu sendiri. Tes itu sendiri minimal harus mampu mengukur pembelajaran atau unit tujuan pembelajaran. Tetapi tidak semua tes dikonstruksi untuk mengukur prestasi, dan suatu kriteria akan lebih relevan dari pada tujuan yang diinginkan pengajar. Misalnya seseorang tidak dapat menentukan apakah tes mampu mengukur kejiwaan dengan menguji isi. Peristiwa empirik dibutuhkan untuk mendemonstrasikan tes secara nyata mengukur hal seperti itu.

Face Validity

Face validity suatu tes adalah tingkat di mana nampak relevan, penting dan menarik bagi peserta ujian. Guru ingin siswa menikmati dalam melakukan tes dan merasakan bahwa mereka tidak menya-nyiakan waktu mereka. Yang penting adalah apakah kenyataannya siswa membuang-buang waktu mereka pada tes yang tidak valid dan reliabel - tidak masalah bagaimana tes nampak relevan. Meskipun begitu, untuk memotivasi dan hubungan yang baik adalah penting bagi siswa dan orang tua merasa tes adalah penting. Keluhan untuk sekolah utama siswa dapat dikurangi jika tes tidak nampak sepele dalam isi atau tidak menggunakan bahasa yang rendah kepada siswa.

Beberapa butir mungkin mempunyai face validity, tapi kurang validitas empirik. Itu mungkin bagi siswa untuk menikmati pengambilan pengalaman tanpa menggunakan tes. Mungkin juga mengukur validitas empiris, tapi mempunyai face validity yang kecil. Sering menulis ulang butir-butir dapat membuat butir nampak lebih relevan. Tes aritmatika dasar untuk siswa SMA yang lambat sebaiknya menggunakan cara dewasa, contoh tes membaca untuk seleksi sekretaris perlu menggunakan contoh praktis dan menghindari kepustakaan, tes matematika dengan ilmu fisika kelas perlu menggunakan contoh dari ilmu fisika.

Harus jelas bahwa face validity adalah tidaklah penting ke bentuk validitas lain. Jika mungkin bagaimanapun butir-butir perlu nampak relevan untuk peserta ujian sebab jika tes nampak sepele dan kekanak-kanakan mungkin, validitas empiris mungkin lemah.

VALIDITAS KRITERIA TERKAIT

Istilah *validitas kriteria terkait* pertama kali dikenalkan pada tahun 1966 edisi *standar tes pendidikan psikologi dan manual* (APA: 1966:13) dan digunakan kembali pada tahun 1974 (APA, 1974:26). Hal ini berarti bahwa korelasi pengukuran dengan kriteria eksternal. Jika pengukuran digunakan untuk memprediksi perilaku dimasa depan (tingkat, ranking, skor pada tes dan lain-lain), maka korelasi seperti disebut *koefisien validitas prediktif*. Jika validitas diestimasi melalui hubungan pengukuran dengan kriteria yang dapat diperoleh saat ini maka korelasi ini disebut *koefisien validitas konkuren*. Validitas konkuren adalah nilai dalam menentukan apakah suatu tes yang dikembangkan dapat diganti dengan yang lebih kompleks atau prosedur penilaian yang mahal.

Validitas Prediktif.

Validitas prediktif digunakan untuk *memprediksi* atau mengestimasi bagaimana individu akan melakukan berbagai tugas berdasarkan *kriteria* yang saling berhubungan. Prediksi menunjukkan perbedaan waktu antara peroleh pengukuran pada tes prediktor

dan kriteria terhadap tes yang divalidasi. *Kriteria* adalah standar yang diterima untuk mengukur sifat atau kemampuan yang diprediksi. Karena minimal nilai rata-rata C dapat diterima disekolah-sekolah maka IP sering digunakan sebagai kriteria atau standar keberhasilan. Keberhasilan dalam pekerjaan dapat diukur melalui peningkatan gaji, rekomendasi pengawas, masa kerja dan lain-lain.

Terdapat ketidaksesuaian tentang kriteria keberhasilan yang digunakan. Ketidakpuasan dengan nilai sebagai kriteria telah menyebabkan beberapa sekolah mengganti kriteria tersebut dengan kehadiran atau partisipasi dalam komunitas terhadap nilai tradisional. Karena setiap prediktor nampaknya berbeda hubungannya dengan setiap kriteria maka kemungkinan koefisien validitas prediktif juga berbeda. Konsekuensinya tidak satupun koefisien validitas prediktif mewakili validitas dari sejumlah pengukuran. koefisien validitas prediktif yang berbeda dapat diperoleh tergantung pada *kriteria* apa yang disetujui dan *prediktor* apa yang dipilih.

Koefisien validitas prediktif adalah hubungan antara skor tes prediktor dan kriteria pengukuran yang diperoleh. Tes prediktor dikonstruksi atau dipilih untuk sampel keterampilan, atribut atau perlakuan diperoleh melalui kriteria. Koefisien validitas prediktif yang tinggi berarti bahwa prediktor mengukur perlakuan yang sama sebagai kriteria, jika koefisien rendah maka hubungan keduanya adalah kecil. Koefisien validitas prediktif memiliki pertimbangan bervariasi tetapi korelasi dari 0,60 atau 0,70 dianggap tinggi.

FAKTOR-FAKTOR YANG MEMPENGARUHI KOEFISIEN VALIDITAS PREDIKTIF

Waktu Antara Pengukuran Prediktor dan Kriteria

Prediksi-prediksi cenderung menjadi kurang akurat dalam jangka waktu panjang (lebih akurat dalam jangka waktu pendek), karena banyak faktor kesempatan bisa mempengaruhi baik predictor maupun kriteria, dalam memproduksi korelasi-korelasi yang rendah. Dalam jangka waktu pendek efek-efek kesempatan menurun dan koefisien validitas prediktif bertambah.

Koefisien validitas prediktif biasanya diperoleh untuk tujuan “seleksi” dan “penempatan”, dan konstruktor sebuah tes mungkin tidak bisa menunggu dalam waktu yang lama untuk menghubungkan pengukuran predictor dan kriteria. Konstruktor-konstruktor tes memilih mengembangkan koefisien validitas prediktif dalam jangka pendek, menyadari bahwa penundaan yang lama mungkin menghasilkan koefisien-koefisien yang rendah.

Keterbatasan praktis yang lain adalah kebutuhan untuk mengetes banyak orang untuk meyakinkan bahwa jumlah yang memadai akan tersedia jika sudah waktunya untuk mengukur kriteria. Sebagai contoh, untuk validasi sebuah ujian masuk SMU, mungkin perlu mengetes ribuan siswa, karena akan banyak yang gagal untuk bersekolah, dimana studi validasi tersebut diadakan, dan yang lain akan *drop-out* sebelum pengukuran kriteria dihasilkan. Karena pengurangan, data kriteria mungkin akan diperoleh pada akhir tahun pertama.

Efek-efek Reliabilitas

Koefisien validitas maksimum berhubungan langsung dengan reliabilitas. Formula yang menggambarkan hubungan ini adalah :

$$\text{Validitas maksimum} = \sqrt{\text{reliabilitas}}$$

Dengan kata lain, jika reliabilitas dari seperangkat pengukuran adalah 0,7 maka validitasnya tidak bisa lebih tinggi dari 0,7 atau 0,837. Mungkin saja suatu tes akan berkorelasi lebih tinggi dengan kriteria eksternal (validitas) daripada dengan dirinya sendiri (reliabilitas).

Variabilitas Kelompok

Di depan telah diperlihatkan bahwa korelasi cenderung meningkat dengan meningkatnya heterogenitas kelompok, dan seseorang mungkin mengharapkan hubungan ini, untuk memperoleh keduanya, baik validitas maupun reliabilitas, karena mereka diekspresikan sebagai hubungan-hubungan. Tetapi karena reliabilitas adalah perbandingan antara varians yang sebenarnya dan varians yang diperoleh, dan varians yang diperoleh adalah jumlah dari varians sebenarnya ditambah varians error, maka reliabilitas mungkin bisa didefinisikan sebagai varians sebenarnya dibagi varians sebenarnya plus varians error.

$$\text{reliability} = \frac{\text{true variance}}{\text{obtained variance}} = \frac{\text{true varian}}{\text{true variance} + \text{error variance}}$$

Peningkatan varians yang diperoleh (dengan penambahan item-item pada tes atau dengan memberikan tes tersebut kepada sampel yang lebih homogen) mempengaruhi reliabilitas, tergantung apakah peningkatan variabilitas berhubungan dengan true variance/ error variance. Jika peningkatan varians yang diperoleh terjadi karena true varians, reliabilitas akan meningkat. Jika itu karena error variance, reliabilitas akan menurun. Jika seseorang mampu menentukan setiap true score siswa, true score- true score tersebut akan reliable sempurna. Jika sebuah angka yang berbeda dan diseleksi secara acak ditambahkan ke setiap true score ini, variabilitas skor mentah akan meningkat, tetapi reliabilitas akan menurun, karena errornya meningkat. Peningkatan variabilitas kelompok akan meningkatkan semua korelasi, termasuk koefisien reliabilitas dan validitas, hanya jika peningkatan variabilitas tersebut disebabkan oleh peningkatan varians true score dan penurunan varians error.

Reliabilitas Kriteria

Kriteria, seperti predictor-prediktor, adalah pengukuran yang bisa salah. Apakah itu rata-rata point kelas, penilaian guru-guru, atau jumlah soal yang terselesaikan dengan benar, semua itu adalah subjek dari kesalahan pengukuran dan ketidak reliabelan. Seorang siswa yang menyelesaikan 50 soal dengan benar hari ini, mungkin hanya 40 soal untuk hari berikutnya. Rata-rata dari skor yang diperoleh siswa masih merupakan perkiraan terbaik terhadap hasil yang sebenarnya, dan segala perbedaan nilai ini dengan hasil yang diperoleh siswa pada tes ekuivalen yang lain adalah kesalahan pengukuran.

Karena kesalahan-kesalahan pengukuran ini adalah acak (karena kesempatan), tak ada variabel yang berkorelasi dengan mereka. Korelasi berasumsi konsistensi dari satu variabel ke variabel berikutnya, dan kesalahan-kesalahan pengukuran, pada dasarnya, tidak konsisten.

Jika kesalahan pengukuran bisa dihilangkan, korelasi yang dihasilkan akan terjadi antara true scores. Kesalahan pengukuran melemahkan koefisien korelasi. Meskipun

kriteria sendiri bisa salah, mereka harus menjadi standar yang sempurna terhadap pengukuran prediktor yang dievaluasi. Skor pada tes prediktor bisa muncul cacat jika kenyataannya kriteria tidak reliabel. Untungnya, ini memungkinkan untuk mengestimasi hubungan antara tes yang bisa salah, dengan kriteria yang reliabel sempurna, melalui teknik statistik. Ini adalah koreksi yang beralasan, karena prediktor tes harusnya tidak dianggap cacat jika kriterianya tidak reliabel.

Dalam melaporkan koefisien-koefisien validitas, penting untuk membedakan antara nilai-nilai yang dikoreksi atau tidak. Koefisien yang tidak dikoreksi selalu ditentukan oleh hubungan sebuah predictor yang bisa salah, dengan sebuah kriteria yang bisa salah. Karena itu mereka berisi kesalahan pengukuran dan bagaimanapun, varians yang sebenarnya adalah sama untuk keduanya. Pada prakteknya, sebuah predictor tidak bebas dari error dan kekurangan reliabilitasnya harus dicerminkan dalam koefisien validitas. Kriteria di sisi lain, haruslah reliabel sempurna dan sangat beralasan untuk mengoreksi kriteria atas atenuasi, sepanjang itu jelas bahwa korelasi adalah nilai teoritis dan dikoreksi, daripada koefisien yang tidak dikoreksi.

Formula untuk *koreksi atenuasi untuk kriteria* adalah :

$$r = \frac{\text{Validity coefficient}}{\text{criterion reliability}}$$

Juga memungkinkan untuk mengoreksi predictor maupun kriteria dari atenuasi. Ini akan menghasilkan sebuah koefisien validitas prediktif yang maksimum, karena efek dari kesalahan pengukuran akan dieliminasi dari keduanya. Tetapi koefisien-koefisien sebaiknya tidak digunakan dalam membuat keputusan-keputusan prediksi praktis berkenaan dengan individu-individu karena mereka mengeliminasi kesalahan pada apa yang sebenarnya diukur.

Formula untuk *koreksi atenuasi bagi keduanya (skor predictor dan kriteria)* adalah :

$$r = \frac{\text{Validity coefficient}}{\text{predictor reliability} \times \text{criterion reliability}}$$

Jumlah Skor Prediktor

Meningkatkan panjang suatu pengujian pada umumnya meningkatkan reliabilitas (lihat hal 270-271) dan validitas, sebab validitas yang maksimum meningkatkan fungsi reliabilitas. Peningkatan validitas akan terjadi jika butir yang ditambahkan dapat diperbandingkan dengan tes yang asli. Selanjutnya, dengan asumsi kriteria dan individu penguji tetap tanpa perubahan.

Menambahkan soal baru pada pengujian akan memungkinkan mempunyai efek lebih besar pada validitas. Seperti dapat dilihat pada table 10.2. Jika 30 butir ujian yang asli mempunyai koefisien reliabilitas 0,700 dan koefisien validitas 0.400, penambahan panjang dari pengujian lebih mempengaruhi reliabilitas dari pada validitas, dan

peningkatan terbesar terjadi koefisien semakin layak. Apalagi asumsi kriteria dan individu tetap.

Table 10.2 Reliabilir and Validity Coefficients for Tests of Various Lengths

Number of Items	Reliability	Predictive Validity
30	.700	.400
60	.824	.433
90	.875	.447
120	.903	.454
150	.921	.459
180	.933	.462
210	.942	.464

Validitas Concurrent

Koefisien validitas konkuren dapat dihitung ketika pengukuran pada prediktor dan kriteria tes diperoleh pada saat yang sama. Jika pengajar ingin mengetahui seberapa baik pengukuran *memperdiksi* kriteria maka menggunakan validitas prediktif adalah tetap. Pada saat yang lain ketika pengajar ingin mengetahui apakah tes tersebut mampu mendiskriminasi antara siswa yang berhasil dan tidak berdasarkan beberapa kriteria tugas. *Koefisien validitas konkuren tinggi* menunjukkan bahwa tes mampu mendiskriminasi individu berdasarkan kriteria keberhasilan seperti mengukur melalui gaji, indeks prestasi dan lain-lain; *koefisien validitas prediktif tinggi* menunjukkan bahwa tes mampu memprediksi individu mana yang akan berhasil.

Satu contoh yang dapat digunakan untuk mengklarifikasi antara konkruen dan validitas prediktif. Misalnya suatu sekolah psikologi ingin mengembangkan tes untuk memprediksi siswa mana yang akan memperoleh perlakuan psikiater. Pertama ahli psikologi memberikan tes pada kelompok besar siswa SMA kemudian menunggu 5 tahun untuk mengukur kesesuaiannya. Ketika penelitian telah lengkap maka ahli psikologi akan melaporkan efektifitas tes dalam *memprediksi* individu mana yang memperoleh perlakuan dalam contoh ini terdapat selang waktu (5 tahun) antara prediktor dan kriteria (jumlah siswa yang memperoleh perawatan psikiater).

Beberapa orang dapat dengan sabar menunggu waktu yang panjang tersebut, namun umumnya mereka ingin tahu apa hubungan *bersamaan* antara skor tes dengan kriteria. Untuk menghindari menunggu 5 tahun maka dapat diberikan tes perlakuan psikiater dan dihubungkan dengan kurangnya gejala emosional. Korelasi antara skor tes (prediktor) dan beberapa kriteria diperoleh pada saat yang sama dapat mengukur validitas konkuren. Skor tinggi pada tes eksperimental harus diasosiasikan dengan tingginya tingkat kesesuaian atau sebaliknya.

Jika tes tidak dapat membedakan antara ketidak sesuaian dan individu normal maka prediksi kriteria ke depan tidak tepat. Koefisien validitas konkruen memberikan nilai maksimum terhadap koefisien validitas prediktif. Jika koefisien validitas konkuren 0,60 maka validitas prediktif akan lebih rendah, asumsinya bahwa kriteria pengukuran sama.

VALIDITAS KONSTRUK

Validitas dapat didefinisikan sebagai pengembangan pengukuran yang digunakan dalam pengambilan keputusan sesuai dengan tujuan yang diberikan. Validitas dari judul tes harus didukung oleh peristiwa dalam tes manual. Misalnya tes motivasi harus

menimbulkan peristiwa yang dapat dilakukan dalam kenyataan untuk mengukur motivasi dan dapat digunakan dalam situasi untuk keputusan yang berhubungan dengan konstruk. Menentukan validitas konstruk adalah penting ketika pengembang tes atau pengguna tes ingin mempelajari lebih banyak tentang kualitas psikologi yang diukur melalui tes dari pada dipelajari melalui kriteria tunggal yang berhubungan dengan koefisien validitas (APA 1974: 30). Misalnya keputusan yang berhubungan dengan keterampilan siswa pada suatu tes yang didesain untuk mengukur kreatifitas, dapat dikembangkan jika guru memahami bagaimana skor tersebut berhubungan dengan konstruk seperti prestasi atau intelegensi. Validitas konstruk juga penting dalam menentukan perlakuan apa atau kemampuan yang dikembangkan dalam pengukuran tes terbaru.

Tujuan lain dalam menentukan validitas konstruk adalah berhubungan dengan pengembangan dan perbaikan teori pendidikan dan psikologi. Observasi empiris (pengukuran) memberikan data input untuk mengklarifikasi dan mendefinisikan teori. Teori yang mengungkap bahwa frustasi menghasilkan agresi harus memiliki beberapa cara dalam mengukur frustasi dan agresi tersebut. Jika perbedaan pengukuran frustasi berhubungan secara positif dan tinggi, hal ini menunjukkan bahwa *frustasi* adalah suatu perlakuan tunggal walaupun diukur dengan pendekatan yang berbeda. Dalam beberapa cara *panjang* dapat diukur dengan instrumen yang berbeda. Seseorang menunjukkan *tidak frustasi* dan juga tidak agresif. Para ahli teori memperoleh informasi sebanyak mungkin untuk memahami makna hipotesis konstruk. Peristiwa akan mendukung atau gagal mendukung konstruk tetapi tidak dapat membuktikan proposisi yang diberikan. Peristiwa lain diperoleh berdasarkan kondisi yang berbeda dan dengan subjek yang berbeda dapat diperoleh beberapa modifikasi konstruk.

Tes kepribadian dan intelegensi secara khusus dilakukan untuk mengukur konstruk teoritik atau perlakuan. Misalnya teori intelektual dapat memberi dukungan jika tes dikembangkan untuk mengukur perilaku apa saja dari teori yang diprediksi. J.P. Guilford (1967) mempunyai hipotesis bahwa intelegensi terdiri dari 120 kecakapan dan dia telah melakukan penelitian yang luas untuk membuktikan hal tersebut. Beberapa peristiwa mempunyai korelasi yang rendah, dan peristiwa lain mencakup perbandingan prestasi siswa sebagai tes prediksi yang diperoleh dari teori Guilford dan berbagai pertimbangan teori lain. Dalam penelitian itu IP dan skor pada tes matematika (aljabar) digunakan sebagai kriteria. Keith Holy dan William Michael (1972) menemukan bahwa tes konstruk berdasarkan teori Guilford merupakan prediktor yang konsisten terhadap dua kriteria tersebut dan menghemat waktu untuk mengadministrasinya. Untuk memperjelas makna intelegensi beberapa penelitian harus dilakukan menggunakan tes dan kriteria lain untuk mengukur aspek-aspek yang berbeda.

Umumnya proses validitas konstruk mencakup minimal 6 langkah :

1. Justifikasi eksplisit bahwa konstruk adalah penting dalam pendidikan atau psikologi. Mengembangkan konstruk yang tidak berhubungan dengan teori atau pengembangan praktek mencakup tujuan tertentu. Untuk menjustifikasi kebutuhan teori intelektual Guilford (1959) telah mendemonstrasikan nilai integrasi teori intelegensi dan menunjukkan implikasi teori tersebut terhadap teori psikologi, tes kejuruan dan praktek pendidikan.
2. Perbedaan harus dibuat antara hipotesis konstruk dan konstruk lain walaupun nampak sama. Misalnya Guilford menyebut "kreatifitas" lebih baik didefinisikan *berpikir difergen*. *Berpikir divergen* bukan berarti mengukur usaha yang dipertahankan tetapi justru lebih mempertimbangkan hal-hal yang logis.

3. Hipotesis konstruk harus terukur. Teori Guilford tentang postulat intelektual bahwa terdapat perbedaan pengukuran antara “bilangan” dan “simbolik” dalam berpikir divergen. Berpikir divergen bilangan membutuhkan konstruksi tes yang mengukur variasi dari perbedaan respons terhadap ukuran, warna, bentuk, lokasi, dan susunan. Sebaliknya berpikir divergen simbolik membutuhkan penggunaan huruf, angka, atau simbol-simbol konvensional lain. Tes-tes yang digunakan untuk mengukur berbagai konstruk hipotesis telah disusun oleh Guilford.
4. Bukti/keterangan haruslah diperoleh dari sumber yang berbeda, untuk mendukung konstruk. Biasanya investigator mencoba mengembangkan tes-tes yang berbeda, yang secara independen mengukur trait yang sama. Sebagai contoh : *Figural divergent thinking* harus bisa diukur oleh tes-tes yang berbeda, yang mempunyai ukuran, warna, bentuk, lokasi dan tekstur yang sama. Jika tes-tes yang berbeda semua mengukur konstruk yang sama, maka konstruk tersebut mempunyai validitas *convergent* (Campbell & Fiske, 1959).
5. Bukti/keterangan haruslah diperoleh untuk memperlihatkan bahwa konstruk tidak berkorelasi dengan faktor-faktor yang tidak relevan. Konstruk harus mempunyai validitas *discriminant*. Sebagai contoh : Sebuah tes atas pemikiran yang berbeda/baru, haruslah tidak berkorelasi dengan kuat/ positif, dengan sebuah pengukuran atas kekakuan, karena kedua trait tersebut secara logis bertentangan. Memperlihatkan apa yang tidak digambarkan sebuah konstruk, sama pentingnya dengan memperlihatkan apa yang digambarkannya.
6. Konstruk dimodifikasi sesuai dengan informasi tambahan. Ketika bukti/ keterangan baru terkumpul, investigator harus memodifikasi sifat-sifat dari konstruk tersebut. Kapan pun konstruk tidak efisien/ gagal memprediksi suatu hipotesa, konstruk tersebut memerlukan modifikasi. Bukti/ keterangan baru akan memodifikasi harapan-harapan, yang pada gilirannya menyarankan kepada investigator, pendekatan-pendekatan baru untuk diikuti.

PENUTUP

Sebuah test adalah valid jika ia mengukur apa yang mau diukur. Tipe utama dari validitas adalah validitas konten, validitas berhubungan kriteria, dan validitas konstruk. Validitas konten didasarkan pada eksaminasi/pemeriksaan subyektif item test. Ada dua tipe validitas konten: validitas face dan validitas logik. Sebuah test memiliki validitas face jika eksaminasi item menghasilkan kesimpulan bahwa item itu mengukur apa yang mereka dimaksudkan untuk mengukur. Kerangka kerja konseptual yang dikembangkan secara mutakhir untuk validitas adalah validitas konstruk. Penggunaan validitas konstruk adalah sebuah proses yang terus menerus yang melibatkan verifikasi prediksi yang dibuat mengenai skor test.

Daftar Pustaka

- Abel, R. L. 1972. *Essential of Educational Measurement* (2rd ed). Englewood Cliffs. New Jersey: Prentice-Hall, Inc.
- Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.

Allen, MJ & Yen WM. 1979. *Introduction to Measurement Theory*. Belmont, California: Wadsworth, Inc

Cronbach, L. J. & Quirk, T. J. (1976). *Test validity*. In *International Encyclopedia of Education*. New York: McGraw-Hill.

Sudjana, N. 1988. *Penilaian hasil proses belajar mengajar*. Bandung: Rosda Karya.

Sumadi Suryabrata. 1987. *Pengembangan Tes Hasil Belajar*. Jakarta: Rajawali Press.

Sax , Gilbert (), *Principles Of Educational and Psychological Measurement and Evaluation*.