

Validitas Instrumen Penelitian Psikologi dan Pendidikan

*Modul Mata Kuliah Instrumen Penelitian
Semester Genap 2019/2020*



Anindito Aditomo, Ph.D.



Program Studi S3 Psikologi
Fakultas Psikologi, Universitas Surabaya

Daftar Isi

Daftar Isi.....	1
Petunjuk penggunaan modul.....	2
Tujuan dan evaluasi pembelajaran.....	3
Makna validitas dalam pengukuran dan penelitian.....	4
Pendekatan pragmatis terhadap validitas.....	6
Daya prediksi terhadap sebuah kriteria.....	6
Konten tes dan konsistensi internal.....	8
Pendekatan teoretis terhadap validitas.....	12
Latihan.....	16
Referensi.....	19

Petunjuk penggunaan modul

Modul ini ditulis sebagai bahan belajar mandiri bagi mahasiswa peserta mata kuliah Instrumen Penelitian, serta mata kuliah lain yang membutuhkan pengetahuan tentang konsep validitas dan bagaimana mengevaluasi validitas dalam konteks pengukuran psikologi dan pendidikan. Modul ini memuat bagian teori, serta beberapa aktivitas yang dapat digunakan untuk menilai pemahaman diri (*self assessment*). Meski tidak diwajibkan, mahasiswa diharap menggunakan modul ini sebelum pertemuan yang membahas topik atau materi yang terkait di kelas. Dengan demikian, waktu pertemuan di kelas dapat digunakan secara lebih optimal untuk kegiatan yang bersifat interaktif.

Tujuan dan evaluasi pembelajaran

Setelah mempelajari dan mengerjakan aktivitas dalam modul ini, mahasiswa diharapkan memiliki pemahaman awal tentang konsep validitas hasil pengukuran menggunakan suatu instrumen dalam konteks penelitian psikologis dan pendidikan.

Pemahaman ini ditandai oleh kemampuan untuk:

1. menjelaskan dua perspektif (pengukuran dan fungsional) dari validitas.
2. menjelaskan dan memberi contoh pendekatan pragmatis terhadap validitas.
3. menjelaskan dan memberi contoh pendekatan teoretis terhadap validitas.
4. menjelaskan dan memberi contoh pendekatan argumentasi terhadap validitas.

Tidak ada penilaian formal untuk mengevaluasi hasil belajar dari modul ini. Meski demikian, dosen akan meminta mahasiswa untuk menjelaskan makna konsep validitas, memberi contoh-contoh baru, serta menceritakan relevansi materi dalam modul ini dengan topik dan metodologi yang akan digunakan dalam disertasi masing-masing.

Makna validitas dalam pengukuran dan penelitian

Dalam konteks pengukuran, validitas adalah soal seberapa yakin para pengguna instrumen percaya bahwa skor yang dihasilkan mencerminkan apa yang hendak diukur serta dapat digunakan untuk keperluan yang diinginkan (AERA, APA, & NCME, 1999). Definisi ini mencakup dua perspektif tentang validitas yang pernah dikemukakan oleh Crobach (1988). Yang pertama adalah perspektif pengukuran, dan yang kedua adalah perspektif fungsional. Dari perspektif pengukuran, validitas berbicara tentang seberapa kita bisa percaya bahwa sebuah skor memang secara akurat dan presisi mencerminkan konstruk yang hendak diukur. Hasil pengukuran juga harus bisa diandalkan untuk menghasilkan skor yang bisa diperbandingkan, meski dilakukan pada waktu dan pada orang yang berbeda. Dengan demikian, skor seorang peserta tes dapat dibandingkan secara adil dengan skor peserta lain.

Dari perspektif fungsional, validitas berbicara tentang kegunaan hasil pengukuran untuk mencapai tujuan-tujuan tertentu. Contoh penggunaan adalah untuk diagnosis, seleksi, penempatan, dan evaluasi intervensi. Sebuah instrumen boleh jadi menghasilkan skor yang berguna untuk diagnosis, namun kurang berguna untuk melakukan seleksi. Dalam pengertian ini, validitas bukanlah sifat dari sebuah instrumen. Validitas adalah sifat yang menggambarkan interpretasi dan penggunaan hasil pengukuran menggunakan suatu instrumen. Dalam kata-kata Kane (2021), validasi diperlukan sebagai dasar untuk menjustifikasi keputusan yang dibuat berdasarkan hasil pengukuran. Dengan kata lain, validasi dilakukan untuk mendapatkan kepercayaan dari pihak-pihak yang

menggunakan atau yang terdampak oleh keputusan yang dibuat berdasarkan hasil suatu pengukuran. Sebagai contoh, tes untuk menyeleksi calon karyawan tentu harus dipercaya oleh departemen SDM perusahaan tersebut. Hasil tes juga perlu dipercaya oleh para peserta tes. Contoh lain, asesmen yang menentukan status akreditasi sebuah sekolah tentu harus dipercaya oleh masyarakat yang menggunakan status akreditasi untuk menilai kualitas sekolah.

Singkat kata, yang bisa dinilai valid atau tidak bukanlah sebuah instrumen. Yang bisa dan perlu dinilai validitasnya adalah penggunaan skor yang dihasilkan dari kegiatan pengukuran dengan instrumen tertentu. Bagaimana cara mengevaluasi validitas ini? Dalam psikologi dan ilmu-ilmu terkait, pemikiran tentang validitas berkembang seiring dengan perkembangan keilmuan dan kompleksitas penggunaan hasil pengukuran.

Pendekatan pragmatis terhadap validitas

Pada awal abad sampai pertengahan abad ke-20, pengujian validitas dilakukan secara pragmatis. Ketika tes terstandar dalam psikologi mulai berkembang, dan pada umumnya dirancang untuk mengukur intelegensi dan aspek lain dari kemampuan individu. Dalam pendekatan pragmatis terhadap validasi, sebuah tes dikatakan valid ketika skor yang dihasilkan berguna untuk mencapai tujuan dibuatnya tes tersebut. Dengan kata lain, validitas mengacu pada kegunaan atau kemanfaatan sebuah tes untuk mencapai sebuah tujuan.

Daya prediksi terhadap sebuah kriteria

Ketika itu penggunaan tes psikologi umumnya terbatas pada konteks program pendidikan dan pekerjaan untuk melakukan seleksi dan penempatan. Dengan demikian, tes yang valid adalah yang hasilnya berguna untuk memilih peserta tes yang paling cocok untuk sebuah program atau pekerjaan. Tes yang valid harus bisa membedakan antara mereka yang akan berprestasi dari mereka yang akan kesulitan dalam sebuah program atau pekerjaan.

Dalam pendekatan pragmatis ini, teori tidak banyak berperan. Makna dari konstruk yang hendak diukur oleh sebuah tes tidak banyak dibahas. Bahkan sampai pertengahan abad ke-20, pendekatan pragmatis ini masih menjadi arus utama. Bab tentang validitas dalam edisi pertama almanak *Educational Measurement* memuat deklarasi bahwa “skor hasil tes tidak mencerminkan kemampuan seseorang. Justru

sebaliknya, yang disebut sebagai kemampuan sebenarnya hanya rangkuman dari tindakan-tindakannya.” (Cureton, 1951, hal. 641, dikutip dari Kane, 2021, hal. 182). Dengan kata lain, alat ukur-lah yang menentukan barang yang diukur. Kemampuan adalah apapun yang diukur oleh tes kemampuan. Inteligensi adalah apapun yang diukur oleh tes inteligensi.

Selain itu, kriteria yang digunakan dalam validasi juga ditentukan secara pragmatis. Yang dijadikan kriteria adalah indikator yang dianggap secara objektif mencerminkan keberhasilan dalam sebuah konteks nyata. Misalnya, nilai rapor, lama belajar, dan kelulusan dapat dianggap sebagai indikator keberhasilan di sekolah. Kecepatan (efisiensi kerja), akurasi (rendahnya eror atau kesalahan), dan volume output pekerjaan dapat dianggap sebagai indikator keberhasilan di tempat kerja. Validitas tes kemudian diukur dari tinggi rendahnya korelasi antara hasil tes dengan indikator-indikator keberhasilan tersebut. Pemilihan indikator pragmatis sebagai kriteria validitas tidak perlu didasarkan pada justifikasi teoretis. Bahkan sebaliknya: tes psikologi dibuat berdasarkan asumsi tentang kemampuan apa yang melandasi keberhasilan di sekolah dan tempat kerja. Dengan kata lain, indikator pragmatis itulah yang menentukan perancangan sebuah tes. (Dalam sejarahnya, tes intelegensi memang dibuat untuk keperluan pragmatis seperti menyortir siswa dan calon prajurit).

Pendekatan pragmatis dalam validasi instrumen ini sampai sekarang masih digunakan. Istilah yang sering digunakan sekarang adalah “validitas prediktif”. Misalnya, sebuah tes seleksi masuk perguruan tinggi akan dianggap valid ketika hasilnya bisa membedakan

antara calon mahasiswa yang lebih mungkin berhasil dari mereka yang lebih mungkin gagal kuliah di program studi tertentu. Pendekatan ini memang berguna dan cukup simpel untuk diterapkan. Yang perlu dilakukan adalah menentukan kriteria pragmatis sesuai tujuan penggunaan instrumen, kemudian mengumpulkan data menggunakan instrumen tersebut dari cukup banyak responden, dilengkapi dengan data tentang kriteria yang dipilih.

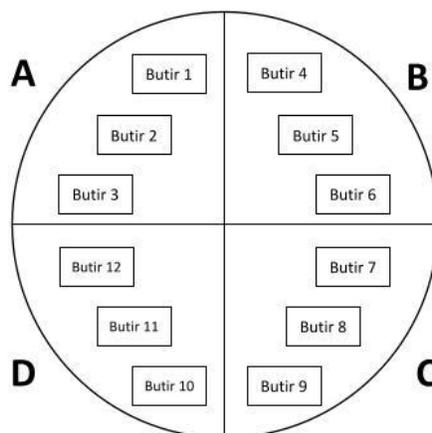
Konten tes dan konsistensi internal

Menilai validitas berdasarkan daya prediksi cocok diterapkan untuk tes yang dipakai dalam seleksi atau penempatan. Hasil tes memang harus bisa memprediksi kinerja di konteks kerja atau program pendidikan. Namun ada kalanya sebuah instrumen dikembangkan bukan untuk memprediksi *outcome* tertentu, melainkan untuk mengukur hasil dari program pembelajaran atau pengalaman terstruktur lain. Contoh yang paling umum adalah tes prestasi yang mengukur penguasaan materi belajar yang telah dipelajari dalam sebuah program. Penguasaan materi tersebut merupakan tujuan dari program, dan belum tentu (tidak harus) bisa memprediksi kriteria lain di luar program yang dijalani. Penguasaan materi berbagai mata pelajaran di sekolah seringkali tidak terkait dengan kemampuan lain di dunia nyata. Karena itu validitas tes semacam ini kurang cocok dinilai berdasarkan daya prediksinya terhadap kriteria lain. Untuk kasus seperti ini, validitas dapat dinilai berdasarkan kelengkapan dan relevansi isi tes.

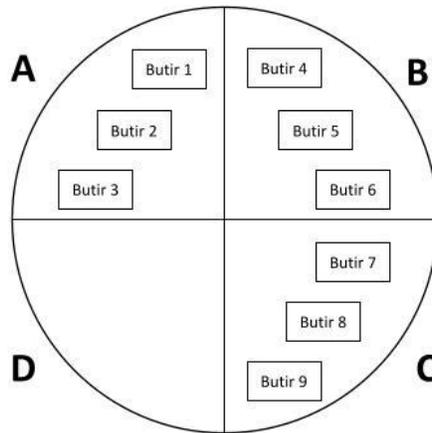
Pada paruh pertama abad ke-20, penilaian terhadap konten tes juga dilakukan secara pragmatis. Artinya, pengembangan tes atau instrumen tidak berangkat dari teori, melainkan dari analisis terhadap domain yang dipandang penting untuk diukur. Dalam

contoh tes prestasi belajar yang dipakai di sekolah, domain yang dianalisis adalah kurikulum yang digunakan atau berlaku. Sebuah tes prestasi dikatakan valid ketika isinya (butir-butir pertanyaannya) mencakup semua materi yang harus dikuasai setelah mengalami suatu program pendidikan. Selain cakupannya harus lengkap, tes tersebut juga tidak boleh memuat butir-butir yang tidak relevan. Dengan kata lain, sebuah tes hanya boleh memuat butir-butir yang relevan dengan domain yang hendak diukur. Sekali lagi, justifikasi terhadap cakupan dan relevansi materi tidak bersifat teoretis. Justifikasinya bersifat pragmatis (misalnya berdasarkan kebijakan tentang kurikulum).

Sebagai ilustrasi, misalkan sebuah tes ingin mengukur prestasi matematika siswa kelas 6. Analisis domain tesnya dilakukan berdasarkan kurikulum matematika kelas 6, yang misalkan mencakup empat sub-domain: (a) pemahaman bilangan, (b) aritmatika, (c) geometri, dan (d) data dan statistika. Masing-masing sub-domain ini harus diukur oleh butir-butir yang memadai.



Jika ada sub-domain yang butirnya kurang atau bahkan kosong (seperti pada gambar di bawah), maka tes tersebut dikatakan kurang valid dari sisi kontennya. Penilaian tentang jumlah butir yang dianggap cukup, serta relevansi setiap butir, dilakukan secara kualitatif oleh orang yang dipandang memiliki otoritas terkait topik yang diukur.



Selain secara kualitatif (melihat relevansi dan kelengkapan), konten instrumen juga dapat dinilai secara kuantitatif menggunakan teknik statistik yang mengukur konsistensi internal. Penilaian konsistensi internal ini dilakukan dengan cara meminta sekumpulan responden untuk mengisi tes atau instrumen. Respon dari sampel tersebut kemudian digunakan untuk menghitung korelasi antar butir, dan korelasi antar butir dengan skor total. Jika korelasi-korelasi tersebut positif dan cukup kuat, maka instrumennya dikatakan memiliki konsistensi internal yang baik. Konsistensi ini mendukung interpretasi bahwa skor total instrumen dapat dianggap mencerminkan satu hal yang sama (entah itu prestasi, kemampuan, atau sifat lain dari individu). Sekali lagi, dalam pendekatan pragmatis, sesuatu yang diukur ini tidak didefinisikan secara teoretis. Yang

penting adalah bahwa butir-butir dalam instrumen secara konsisten mengarah pada satu hal yang sama (Cureton, 1951).

Sebagai ringkasan, pemeriksaan terhadap konten dan konsistensi internal secara pragmatis ini dapat diterapkan untuk menilai validitas instrumen yang mengukur prestasi, keterampilan, kemampuan kognitif, dan atribut kemampuan lain dari individu. Penilaian terkait konten ini melengkapi pemeriksaan terhadap daya prediksi terhadap sebuah kriteria eksternal yang dibahas sebelumnya. Keuntungan dari pemeriksaan terhadap konten adalah kita tidak perlu menunggu sampai bisa mendapatkan data yang mencerminkan kriteria eksternal (output kinerja, misalnya) yang baru bisa diperoleh di masa depan. Kedua jenis pemeriksaan ini sama-sama pragmatis dalam arti tidak berlandaskan pada teori.

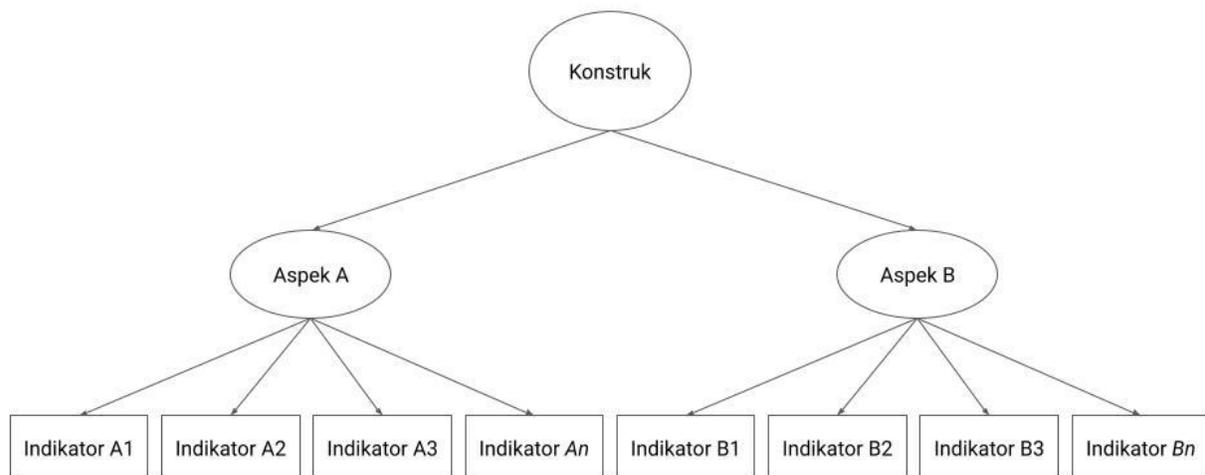
Pendekatan teoretis terhadap validitas

Meski berguna dan mudah dipahami, pendekatan pragmatis dalam validasi menyisakan persoalan. Banyak tema penelitian psikologi dan pendidikan yang tidak memiliki kriteria eksternal yang jelas. Misalnya, ciri-ciri kepribadian seperti sifat ekstrovert, kegigihan, dan perfeksionisme tidak selalu memiliki kaitan yang tegas dengan kinerja atau kriteria eksternal lain. Seseorang yang perfeksionis bisa berprestasi tinggi dalam konteks dan situasi tertentu, namun bisa sangat buruk prestasinya di konteks lain. Kondisi-kondisi afektif seperti rasa syukur, kebahagiaan, dan depresi juga tidak punya kaitan dengan kriteria eksternal yang jelas untuk diprediksi. Apa yang tercakup dalam konsep-konsep seperti ciri kepribadian, kondisi afektif, dan berbagai variabel psikologis lain sulit dianalisis hanya berdasarkan pertimbangan praktis. Karena itu pendekatan pragmatis terhadap validitas tidak memadai.

Sejak pertengahan abad ke-20, para ahli mulai mengemukakan pandangan bahwa validasi hasil pengukuran harus didasarkan pada teori yang jelas. Dalam sebuah makalah klasik, Cronbach dan Meehl (1955) mengajukan istilah “validitas konstruk”, di mana “konstruk” mengacu pada definisi konseptual dari apa yang hendak diukur oleh sebuah instrumen.

Dalam pendekatan yang lebih teoretis ini, sasaran pengukuran harus didefinisikan secara konseptual berdasarkan teori. Justifikasi pragmatis dan empiris semata tidak cukup. Kejelasan teoretis ini penting mengingat keberadaan atribut psikologis memang

bersifat abstrak dan konseptual, alias tidak memiliki padanan objektif dalam realitas. Karena itu, apa yang dimaksud dengan perfeksionisme, atau kebahagiaan, atau atribut lain, akan tergantung pada perspektif teoretis yang diambil oleh pembuat instrumen. Apa saja aspek atau komponen esensial dari perfeksionisme atau kebahagiaan, serta makna dari tiap komponen tersebut, harus didefinisikan secara eksplisit. Kemudian, tiap aspek atau komponen diturunkan ke dalam indikator-indikator yang diasumsikan (secara teoretis) merupakan manifestasi tampaknya. Atributnya sendiri selalu dianggap laten alias tidak terlihat secara langsung.



Untuk menggambarkan relasi antara atribut laten dan indikatornya, para peneliti kerap menggunakan model visual seperti pada gambar berikut. Atribut laten dilambangkan dalam bentuk oval, sedangkan indikator tampaknya dilambangkan sebagai kotak. Anak panah menunjukkan relasi antar elemen. Anak panah (satu arah) dari elemen oval ke elemen kotak menunjukkan relasi sebab-akibat antara atribut dan indikatornya. Maknanya adalah sebuah tinggi rendahnya atribut laten menyebabkan tinggi rendahnya respon pada indikator-indikatornya. Sebagai contoh, responden yang memiliki resiliensi

tinggi akan lebih mungkin menjawab “setuju” (mendapat skor tinggi) pada butir-butir atau problem yang memerlukan resiliensi.

Dalam contoh model di atas, indikator-indikator Aspek A tidak saling berhubungan langsung. Demikian juga untuk indikator-indikator Aspek B. Setiap indikator tersebut memiliki hubungan tidak langsung dengan indikator lain, yaitu melalui aspek latennya masing-masing. Dengan demikian, korelasi antar indikator dianggap nol, karena hubungan antar indikator tersebut terjadi melalui aspek latennya. Demikian juga dengan Aspek A dan Aspek B yang tidak memiliki hubungan langsung. Kedua aspek ini berhubungan secara tidak langsung melalui Atribut yang diukur. Dengan kata lain, dalam model tersebut, korelasi antara Aspek A dan Aspek B dianggap nol. Semua asumsi teoretis ini dapat diuji secara empiris menggunakan teknik statistik seperti uji korelasi dan analisis faktor. Hasil pengujian secara statistik ini mencerminkan struktur internal sebuah instrumen (termasuk, tapi tidak hanya, konsistensi internalnya).



Selain lebih mendasarkan diri pada teori, para peneliti kontemporer melihat validitas sebagai sebuah kesatuan (AERA, APA, & NCME, 1999, 2014). Yang tadinya disebut sebagai "validitas prediktif", "validitas isi", dan jenis-jenis validitas lain sekarang dipandang sebagai bukti-bukti yang melandasi klaim tentang validitas konstruk.

Latihan

1. Baca artikel Aditomo (2019) yang berjudul “Disposisi Berpikir Terbuka secara Aktif: Definisi, Pengukuran, dan Kaitannya dengan Prestasi Akademik.”
 - a. Apa atribut yang hendak diukur oleh instrumen yang dibahas?
 - b. Apakah atribut tersebut didefinisikan berdasarkan teori tertentu? Teori apa yang dipakai?
 - c. Menurut teori yang digunakan, apa saja komponen atau aspek dari atribut yang diukur?
 - d. Jenis bukti validitas apa saja yang digunakan untuk mengevaluasi validitas instrumen tersebut?
 - e. Apakah instrumen tersebut dapat menghasilkan skor yang valid, menurut hasil evaluasi dalam artikel tersebut?

2. Baca artikel Aditomo dan Koehler (2020) yang berjudul “Do student ratings provide reliable and valid information about teaching quality at the school level? Evaluating measures of science teaching in PISA.”
 - a. Apa atribut yang hendak diukur oleh instrumen yang dibahas?
 - b. Apakah atribut tersebut didefinisikan berdasarkan teori tertentu? Teori apa yang dipakai?
 - c. Menurut teori yang digunakan, apa saja komponen atau aspek dari atribut yang diukur?
 - d. Jenis bukti validitas apa saja yang digunakan untuk mengevaluasi validitas instrumen tersebut?

- e. Apakah instrumen tersebut dapat menghasilkan skor yang valid, menurut hasil evaluasi dalam artikel tersebut?
3. Cari dan baca artikel jurnal nasional (kategori Sinta 2 atau Sinta 3) yang melaporkan hasil evaluasi validitas sebuah instrumen di bidang psikologi atau pendidikan.
 - a. Apa atribut yang hendak diukur oleh instrumen yang dibahas?
 - b. Apakah atribut tersebut didefinisikan berdasarkan teori tertentu? Teori apa yang dipakai?
 - c. Menurut teori yang digunakan, apa saja komponen atau aspek dari atribut yang diukur?
 - d. Jenis bukti validitas apa saja yang digunakan untuk mengevaluasi validitas instrumen tersebut?
 - e. Apakah instrumen tersebut dapat menghasilkan skor yang valid, menurut hasil evaluasi dalam artikel tersebut?
4. Cari dan baca artikel jurnal internasional (Q1 atau Q2 di *database* Scopus) yang melaporkan hasil evaluasi validitas sebuah instrumen di bidang psikologi atau pendidikan.
 - a. Apa atribut yang hendak diukur oleh instrumen yang dibahas?
 - b. Apakah atribut tersebut didefinisikan berdasarkan teori tertentu? Teori apa yang dipakai?
 - c. Menurut teori yang digunakan, apa saja komponen atau aspek dari atribut yang diukur?

- d. Jenis bukti validitas apa saja yang digunakan untuk mengevaluasi validitas instrumen tersebut?
- e. Apakah instrumen tersebut dapat menghasilkan skor yang valid, menurut hasil evaluasi dalam artikel tersebut?

Referensi

Aditomo, A. (2019). Disposisi Berpikir Terbuka secara Aktif: Definisi, Pengukuran, dan Kaitannya dengan Prestasi Akademik. *Jurnal Psikologi Teori Dan Terapan*, 10(1), 1-14.

Aditomo, A., & Köhler, C. (2020). Do student ratings provide reliable and valid information about teaching quality at the school level? Evaluating measures of science teaching in PISA 2015. *Educational Assessment, Evaluation and Accountability*, 32(3), 275-310.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anastasi, A. (1950). The concept of validity in the interpretation of test scores. *Educational and Psychological Measurement*, 10(1), 67–78

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.

Cureton, E. E. (1951). Validity. In E.F. Lingquist (Ed.), *Educational measurement*. Washington, DC: American Council on Education.

Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527–535.

Kane, M. (2013). Validating the Interpretations and Uses of Assessment Scores. *Journal of Educational Measurement*, 50, 1–73.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*, 3rd ed. (pp. 13– 103) New York, NY: American Council on Education and Macmillan.