

Kajian Artikel:

BUKTI KESAHIHAN DAN KEANDALAN ALAT UKUR: TANGGAPAN ATAS ARTIKEL “TES KETERAMPILAN OLAH RAGA JUDO BAGI MAHASISWA”

Oleh:

Djemari Mardapi

Universitas Negeri Yogyakarta

Abstract

For any research, the quality of data collected by any instrument, either test or nontest, determines the meaningfulness of the research finding. Therefore, the instrument used in any research should have evidences to indicate that error of measurement of the collected data is small. Two kinds of evidence should be reported in any research, namely the validity and ability of the instrument. The evidences can be obtained by item analysis. The appropriate method of item analysis depends on the reference used in developing the instrument, whether norm reference or criterion reference. All evidences should be interpreted based on the meaningfulness of the indices, not based on the statistical significance test, because the statistical significance test depends on the sample size.

key words: validity, reliability, item analysis, meaningfulness, and significance test

Pendahuluan

Salah satu ciri penelitian dengan pendekatan kuantitatif adalah menggunakan data numerik yang diperoleh dari hasil pengukuran. Semua peneliti berusaha agar data hasil pengukuran

merupakan konstruk yang ingin diukur dan mengandung kesalahan sekecil mungkin. Untuk itu alat ukur yang digunakan harus memiliki bukti kesahihan (*validity*) dan bukti keandalan (*reliability*).

Konsep kesahihan menunjuk pada ketepatan, keberartian, dan kegunaan pemberlakuan yang spesifik dari hasil tes. (Standard, 1985). Kesahihan merupakan konsep yang satu, bahkan banyak tipe bukti kesahihan yang berbeda bisa memberikan kontribusi terhadap kesahihan penilaian (Standards, 2000). Kesahihan tidak hanya berdasarkan pada satu studi saja, atau satu bukti saja, tetapi berdasarkan pada kajian yang terus menerus, sehingga diperoleh sejumlah bukti kesahihan suatu alat ukur. Kesahihan alat ukur pada dasarnya terletak pada penggunaan hasil pengukuran yang berupa skor. Secara tradisional, kesahihan alat ukur dikategorikan menjadi tiga, yaitu kesahihan isi, kesahihan kriteria, dan kesahihan *construct* (Standards, 1985). Oleh karena itu semua alat ukur yang digunakan harus memiliki bukti tiga kategori kesahihan tersebut.

Ada dua syarat utama yang harus dipenuhi oleh alat ukur. Syarat pertama, alat ukur yang digunakan harus berdasarkan pada *construct* yang ingin diukur. *Construct* ini diperoleh dari satu atau lebih teori yang digunakan, yaitu yang berupa definisi konseptual. Selanjutnya definisi konseptual dijabarkan menjadi definisi yang operasional dan diuraikan menjadi sejumlah indikator. Pada umumnya *construct* yang ingin diukur oleh sebuah alat ukur adalah satu, sehingga semua skor butir dapat dijumlahkan menjadi skor suatu ubahan, atau dengan kata lain dimensi yang diukur adalah satu. Walau demikian dimungkinkan dimensi yang diukur lebih dari satu, sesuai dengan perencanaan alat ukur. Oleh karena itu semua alat ukur harus memiliki bukti kesahihan, dan pengumpulan bukti ini merupakan proses yang berlangsung terus menerus.

Syarat kedua adalah alat ukur yang digunakan menghasilkan data dengan kesalahan pengukuran sekecil mungkin. Kesalahan

pengukuran ditentukan oleh besarnya indeks keandalan alat ukur. Semakin besar indeks keandalan suatu alat ukur akan semakin kecil kesalahan pengukuran. Semua peneliti berusaha agar alat ukur yang digunakan memiliki indeks keandalan yang tinggi, dan besarnya indeks ini harus dicantumkan pada laporan penelitian.

Untuk memperoleh bukti kesahihan dan keandalan, semua alat ukur harus ditelaah. Telaah terhadap alat ukur berdasarkan pada konsep yang diukur dan teknik penulisan instrumen, termasuk tata bahasa yang digunakan. Telaah awal yang dilakukan sebelum alat ukur diujicoba di lapangan bertujuan untuk mengetahui *construct* teoretis alat ukur dan bahasa yang digunakan pada alat ukur. Telaah kedua dilakukan berdasarkan data hasil ujicoba alat ukur di lapangan. Telaah kedua ini sering disebut analisis butir dan analisis tes, baik untuk alat ukur yang berbentuk tes maupun yang bukan tes. Teknik analisis alat ukur tergantung pada acuan yang digunakan, yaitu acuan norma atau acuan kriteria. Masing-masing menggunakan asumsi yang berbeda, sehingga teknik analisis yang digunakan tidak semuanya sama.

Tulisan ini dibuat untuk menjelaskan secara singkat persyaratan alat ukur yang baik, dan cara melakukan analisis alat ukur, berkaitan dengan artikel Dr. Jumhan Pida sebagai hasil penelitiannya tentang pengembangan tes keterampilan olah raga. Analisis ini dilakukan berdasarkan kajian teoretis dari sejumlah tulisan tentang teori pengukuran.

Kesahihan

Semua peneliti berusaha untuk memperoleh hasil pengukuran yang akurat ditinjau dari segi *construct* dan dari segi kesalahan pengukuran. Untuk itu alat ukur yang digunakan harus memiliki bukti kesahihan dan keandalan. Bukti ini harus diperoleh melalui

kajian teoretis dan kajian empirik. Untuk itu tulisan ini ingin menyajikan cara untuk memperoleh bukti teoretis dan empiris dari suatu alat ukur.

Cara yang digunakan oleh sejumlah peneliti dalam mengumpulkan bukti kesahihan dan keandalan belum berdasarkan pada konsep pengukuran. Sebagai contoh, beberapa peneliti dalam usahanya untuk memperoleh bukti kesahihan dan keandalan alat ukur selalu menggunakan istilah uji kesahihan dan uji keandalan alat ukur. Istilah uji dalam bidang statistik memiliki makna menggunakan teori peluang dalam menguji alat ukur tersebut. Signifikan tidaknya indeks keandalan alat ukur selalu berkaitan dengan besarnya peluang kesalahan.

Hasil uji alat ukur dengan statistik ditentukan oleh besar derajat kebebasannya. Semakin besar sampel yang digunakan akan semakin besar peluang terjadinya signifikansi atau penolakan hipotesis nihil. Makna signifikan ada dua, yaitu signifikan berdasarkan teori peluang dan signifikan dalam makna kemanfaatannya. Pada bidang pengukuran, bukti kesahihan alat ukur ditekankan pada signifikan dalam makna kemanfaatannya, bukan signifikan menurut teori peluang. Hal ini bisa dipahami karena signifikan secara statistik belum tentu signifikan dalam makna kemanfaatannya. Signifikan menurut statistika ditentukan oleh ukuran sampel penelitian atau ujicoba alat ukur. Semakin besar ukuran sampel akan semakin besar pula peluang menolak hipotesis nihil.

Pada dasarnya kesahihan alat ukur terletak pada skor hasil pengukuran, yaitu seberapa jauh skor yang diperoleh itu menunjukkan besarnya ubahan yang diukur pada suatu objek tertentu. Bukti kesahihan dilihat pada skor hasil pengukuran. Kesahihan alat ukur dikategorikan menjadi tiga yaitu kesahihan *construct*, kesahihan isi, dan kesahihan kriteria. Ketiga kesahihan itu memiliki hubungan yang erat. Kesahihan *construct* berhubungan dengan kesahihan isi

alat ukur. Demikian pula kesahihan kriteria memiliki hubungan dengan kesahihan isi. Walau ketiga kesahihan itu saling terkait, namun semuanya harus memberikan bukti empirik.

Bukti kesahihan *construct* diperlukan bila pengguna tes ingin membuat inferensi dari skor tes ke unjuk kerja yang dapat dikelompokkan di bawah label *construct* psikologi (Crickler & Algina, 1986). Kesahihan isi diperlukan untuk menilai apakah butir-butir tes atau instrumen mewakili *construct* yang ingin diukur. Dengan kata lain, kesahihan isi suatu alat ukur menunjukkan derajat keterwakilan ranah yang diukur oleh sampel butir alat ukur itu. Kesahihan kriteria dikategorikan menjadi dua, yaitu kesahihan konkuren dan kesahihan prediktif. Bukti kesahihan konkuren alat ukur diperoleh dengan mencari kuatnya hubungan atau besarnya koefisien korelasi antara skor hasil pengukuran instrumen yang dibuat oleh peneliti atau pengembang alat ukur dan hasil pengukuran dengan alat ukur yang baku, untuk objek pengukuran yang sama. Dengan kata lain, kedua alat ukur itu diberikan pada objek yang sama dan hasilnya dikorelasikan. Semakin tinggi korelasi akan semakin baik alat ukur tersebut. Kesahihan prediktif merupakan korelasi antara skor dari suatu tes untuk seleksi, misalnya keberhasilan belajar, dengan skor kriteria yang terjadi pada waktu mendatang, misalnya indeks prestasi.

Bukti kesahihan *construct* secara teoretis dilihat dari teori yang digunakan, serta kisi-kisi dan indikator yang ditentukan. Bukti empirik kesahihan *construct* dilihat dari hasil analisis faktor, baik yang eksploratori maupun yang konfirmatori. Menurut Reymen & Joreskog (1993) analisis faktor adalah istilah yang generik yang digunakan untuk menjelaskan sejumlah metode yang dirancang untuk menganalisis *interrelationship* antar ubahan yang menghasilkan *construct* dari ubahan hipotetis. Menurut Nunally (1978) kesahihan *construct* adalah kesahihan faktorial. Kesahihan ini

menjawab pertanyaan apakah alat ukur mengukur yang ingin diukur. Menurut Heppner, Kivlighan, dan Wampold (1992) definisi kesahihan *construct* terfokus pada seberapa baik ubahan yang dipilih oleh peneliti menyatakan *construct* hipotetis. Secara sederhana dapat dikatakan bahwa analisis factor membantu menjelaskan seperangkat data dengan melakukan reduksi data menjadi sejumlah faktor yang lebih sedikit. Dengan kata lain, analisis faktor membantu menjelaskan data dengan melakukan reduksi sejumlah data ke bentuk atau ukuran yang bisa dikelola.

Ada dua teknik analisis faktor untuk menunjukkan bukti kesahihan konstruk, yaitu analisis faktor eksploratori dan analisis faktor konfirmatori. Analisis faktor eksploratori memiliki tujuan untuk menghasilkan teori dengan sifat-sifat dasar literatur yang lemah. Prosedurnya adalah: (a) menentukan jumlah faktor, (b) menentukan apakah faktor-faktor berkorelasi atau tidak, (c) menentukan beban semua faktor yang ditimbulkan oleh semua ubahan bebas atau butir soal.

Analisis butir dilakukan untuk mengetahui besarnya tingkat kesulitan, daya beda, efektivitas pengecoh, untuk tes bentuk pilihan ganda. Tingkat kesulitan didefinisikan sebagai perbandingan antara banyaknya peserta tes yang menjawab benar dan banyaknya seluruh peserta tes. Besarnya tingkat kesulitan yang diterima adalah dari 0,30 sampai 0,80. Daya beda ditunjukkan oleh besarnya korelasi antara skor butir dan total skor. Besarnya daya beda yang diterima adalah 0,30 atau lebih besar. Efektivitas pengecoh dilihat pada jumlah jawaban salah yang dipilih oleh peserta tes. Dikatakan efektif jika dipilih oleh paling sedikit 5% dari seluruh peserta tes.

Keandalan

Teknik statistik dalam mengolah data tidak dapat mengubah data yang tidak baik menjadi baik. Kualitas hasil analisis data tergantung pada kualitas data yang diperoleh. Beberapa teknik statistik, seperti analisis regresi, menggunakan asumsi bahwa data pada ubahan bebas tidak memiliki kesalahan pengukuran (Pedhazur, 1982). Besarnya kesalahan pengukuran dapat dilihat dari besarnya indeks keandalan alat ukur menurut teori tes klasik, dan besarnya fungsi informasi butir dan fungsi informasi tes, menurut teori respons butir.

Ketepatan data hasil pengukuran dipengaruhi oleh kualitas alat ukur, cara mengukur, dan kondisi objek yang diukur. Kesalahan dalam pengukuran pada dasarnya digolongkan menjadi dua, yaitu kesalahan yang bersifat sistematis dan yang bersifat acak. Kesalahan sistematis terjadi apabila hasil pengukuran lebih besar atau lebih kecil dari kemampuan yang sebenarnya. Misalnya ada dosen atau guru yang pemurah dan hemat dalam memberi nilai. Apabila sifat pemurah dan hemat ini berlaku pada semua peserta tes, maka hasil tes ini dikatakan memiliki kesalahan sistematis. Selain itu kesalahan sistematis bisa terletak pada soal dan peserta tes. Ada yang cenderung membuat tes yang mudah, dan ada orang yang cenderung membuat tes yang sulit. Apabila murah dan hematnya dalam memberi skor tidak berlaku pada semua orang tetapi hanya pada beberapa peserta tes saja, maka tes tersebut dikatakan bias.

Khusus mengenai artikel Dr. Jumhan Pida yang berjudul "Tes Keterampilan Olahraga Judo bagi Mahasiswa", dapat dikemukakan hal-hal sebagai berikut:

Dilihat dari judul dan tujuannya, penelitian ini cenderung pada pengembangan tes keterampilan dengan acuan kriteria. Hasil tes ini sebaiknya dikategorikan menjadi dua yaitu bisa dan tidak,

atau lulus dan belum lulus. Asumsi yang digunakan pada tes dengan acuan kriteria adalah bahwa hampir semua orang bisa belajar apa saja, tetapi waktunya mungkin berbeda, sehingga perlu ada program remedial dan program pengayaan. Acuan norma menggunakan asumsi bahwa kemampuan orang itu berbeda, sehingga tes yang baik harus dapat membedakan kemampuan peserta tes.

Mengenai hasil penelitian yang dilaporkan dapat dijelaskan sebagai berikut.

1. Uji objektivitas penilaian antarpakar dinyatakan dengan koefisien korelasi sebesar 0,952 dan 0,910 dengan $p < 0,05$. Koefisien korelasi ini tergolong sangat tinggi, namun tidak perlu diuji signifikansinya secara statistik, karena uji dengan statistik sangat dipengaruhi oleh besar sampel yang digunakan, dengan pengertian bahwa semakin besar sampel yang digunakan akan semakin besar peluang untuk signifikan secara statistik. Penafsiran besarnya indeks ini cukup menggunakan signifikansi dalam makna kemanfaatan.

2. Indeks kesahihan

a. Indeks kesahihan butir

Dalam sejumlah buku induk tentang teori pengukuran istilah indeks kesahihan butir menyatakan koefisien korelasi antara skor butir dan skor kriteria yang berasal dari luar, bukan dari tes itu sendiri (Gulliksen, 1967). Korelasi antara skor butir dan skor total ditafsirkan sebagai besarnya daya beda, bukan validitas butir, dan tidak perlu diuji dengan statistik karena yang penting adalah signifikan dalam arti kemanfaatan, bukan secara statistik, karena signifikan secara statistik

tergantungan pada ukuran sampelnya. Dilihat dari besarnya koefisien korelasi antara skor butir dan skor total, tes yang dikembangkan ini memiliki daya beda yang cukup tinggi, batas minimumnya adalah 0,30.

- b. Indeks kesahihan faktor menyatakan ukuran kuatnya hubungan antara Faktor I dan *construct*. Menurut Nunally (1978), kesahihan *construct* adalah kesahihan faktorial, sehingga perlu dilakukan analisis faktor, eksploratori atau konfirmatori. Oleh karena itu indeks kesahihan faktor yang digunakan dalam artikel tersebut bukanlah bukti kesahihan *construct* seperti yang sering digunakan dalam pengembangan tes, tetapi hanya gambaran tentang homogenitas faktor.
 - c. Demikian pula indeks kesahihan baterai tes yang besarnya 0,477 bukan merupakan bukti kesahihan *construct*, karena kesahihan *construct* ditunjukkan oleh beban faktor untuk masing-masing butir tes dari hasil analisis faktor.
3. Indeks keandalan butir-butir tes, ditunjukkan dengan koefisien korelasi antara tes I dan tes II. Istilah indeks keandalan butir jarang digunakan, karena yang umum adalah indeks keandalan tes, yang terdiri dari sejumlah butir soal. Istilah tes I dan tes II juga kurang jelas, mungkin bukan tes, melainkan butir tes. Istilah tes I dan tes II bisa ditafsirkan dua buah tes yang paralel atau bentuk alternatif.

4. Indeks Keandalan Bateri Tes

Penggunaan rumus Alpha dari Cronbach sudah tepat, tetapi perlu diketahui bahwa rumus ini menggunakan asumsi model pengukuran Tau-ekivalen (Lord & Novick, 1968). Besarnya indeks keandalan tes ini 0,888. Sekali lagi besarnya indeks ini tidak perlu diuji signifikansinya secara statistik, karena yang penting adalah makna kemanfaatannya, dilihat dari besarnya kesalahan pengukuran yang diperkenankan. Dalam banyak buku teori pengukuran dituliskan besarnya indeks ini yang dapat diterima adalah 0,70 atau lebih besar (Linn, 1989).

5. Pengukuran kuatnya hubungan skor antarbutir merupakan prosedur awal dalam melakukan analisis faktor. Besarnya korelasi ini menyatakan seberapa jauh butir tes ini homogen, namun belum bisa memberikan bukti kesahihan *construct* alat ukur. Homogenitas merupakan salah satu syarat bahwa dimensi yang diukur adalah satu, yaitu keterampilan olahraga judo. Oleh karena itu hasil penghitungan ukuran kuatnya hubungan antarbutir belum bisa digunakan untuk menyimpulkan konstruk yang diukur atau jumlah dimensi yang diukur.
6. Penggunaan skor baku T juga bisa dipertanyakan. Seperti telah disebutkan di muka, tes ini lebih tepat menggunakan acuan kriteria bukan acuan norma. Penggunaan skor baku T merupakan bagian dari penggunaan acuan norma. Yang penting, prestasi orang berlatih judo tidak dibandingkan dengan orang lain, tetapi dengan standar yang sudah ditetapkan lebih dahulu. Namun, untuk pertandingan bisa digunakan acuan norma,

karena tujuannya untuk menentukan yang kalah dan yang menang.

Saran

Berdasarkan pembahasan di atas maka disarankan:

1. Analisis butir tes dilakukan dengan menggunakan acuan kriteria, karena masalahnya adalah tes keterampilan.
2. Indeks keandalan tes yang dinyatakan dengan koefisien alpha dari Cronbach tidak perlu diuji signifikansinya secara statistik, cukup dilihat makna kemanfaatannya.
3. Untuk mengetahui kesahihan konstruk suatu instrumen, alat ukur, atau tes, dilakukan analisis faktor, eksploratori atau konfirmatori, tetapi menggunakan sampel yang besar.

Daftar Pustaka

- Gulliksen, H. (1967). *Theory of mental test*. New York: John Wiley & Sons.
- Heppner, P. P., Kivlighan, Jr., D. M & Wampold, B. E. (1992). *Research design in counseling*. Pasific Grove, CA: Brooks/Cole.
- Linn. R. L (Ed.), (1989). *Educational measurement*. New York: Macmillan.

Kajian Artikel Bukti Kesahihan Dan Keandalan Alat Ukur: Tanggapan Atas Artikel "Tes Keterampilan Olah Raga Judo Bagi Mahasiswa"

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. New York: Addison-Wesley.

Nunally, L. C. (1969). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Standards for educational and psychological testing. (1985). Washington, DC: APA.

Standards for quality and fairness. (2000). Princeton, NJ: Educational Testing Services.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.