

**PENGEMBANGAN PENGUJIAN VALIDITAS ISI DAN VALIDITAS KONSTRAK:
INTERPRETASI HASIL PENGUJIAN VALIDITAS**

Miftahun Ni'mah Suseno

Program Studi Psikologi Fakultas Ilmu Sosial dan Humaniora
UIN Sunan Kalijaga Yogyakarta
Jl. Marsda Adisucipto Sleman Yogyakarta
miftahsuseno@gmail.com

Abstract. Validitas bukan merupakan bagian dari instrumen pengukuran, namun lebih melekat pada interpretasi serta penggunaan skor yang dihasilkan tes pada subjek yang relevan. Dapat juga dikatakan bahwa validitas dibuktikan dengan diperolehnya kecocokan empirik antara interpretasi skor dan penggunaannya, serta terkait pula dengan dasar teoritik atribut tujuan ukur sebagai rasionalisasi. Pengukuran merupakan kegiatan yang lazim, bahkan rutin dilakukan dalam dunia psikologi dan pendidikan. Banyak penelitian yang menggunakan validitas aitem sebagai justifikasi valid tidaknya skor yang dihasilkan oleh alat ukur yang digunakan dalam penelitian dalam bentuk korelasi aitem-total atau korelasi aitem-total terkoreksi (*corrected item-total correlation*). Namun persoalan validitas tidak sesederhana itu, namun proses validasinya dilakukan terhadap hasil ukur sehingga bisa membuktikan konstruk yang dikembangkan betul-betul berlaku pada subjek yang menjadi tujuan ukur. Proses pengujian validitas isi salah satunya dapat dilakukan dengan menggunakan *Content Validity Ratio (CVR)* dan pengujian validitas konstruk dapat dianalisis dengan metode *Explanatory Factor Analysis (EFA)* dan *Confirmatory Factor Analysis (CFA)*. Namun hal yang lebih penting bukanlah proses pengujian validitas tetapi lebih kepada interpretasi skor hasil validitas. Validitas adalah ringkasan evaluatif baik dalam bentuk bukti atau konsekuensi interpretasi dan penggunaan skor hasil tes. Pada saat ini yang lebih diutamakan adalah interpretasi skor hasil tes berdasarkan hasil pengujian validitas.

Kata kunci: *validitas isi, validitas konstruk, interpretasi hasil pengujian validitas*

A. Pendahuluan

Dalam penelitian sosial dengan pendekatan kuantitatif ataupun evaluasi pendidikan seringkali kita mengandalkan proses pengukuran dalam proses pengumpulan data maupun melakukan evaluasi terhadap proses pembelajaran. Pengukuran bukan berarti berdiri sendiri begitu saja, tetapi ada bagian yang sangat penting mendahuluinya yaitu tersedianya instrumen atau alat ukur atau alat tes. Instrumen atau tes ini yang akan memberikan gambaran sejauh mana tingkat hasil pengukuran.

Sebelum melakukan pengukuran, hal yang perlu diperhatikan dan diketahui terlebih dulu adalah tingkat validitas dan reliabilitas instrumen. Validitas atau derajat kesahihan sangat penting untuk didalami. Evaluasi terhadap validitas terkait dengan

pertanyaan "Apakah tes tersebut mampu mengungkap atribut yang menjadi tujuan ukur?"

Hal ini didasarkan apakah suatu instrumen atau alat ukur sesuai dengan apa yang mau diukur. Banyak individu atau kelompok serampangan melakukan tes, padahal belum tentu tes itu tepat dengan apa yang dikehendaki. Validitas memiliki makna sejauh mana ketepatan ukur atau kecermatan alat ukur dalam mengukur atribut yang menjadi tujuannya.

Pemahaman yang selama ini berkembang mengenai validitas sering dipahami sebagai suatu ukuran valid atau tidaknya instrumen pengukuran, padahal sebenarnya validitas bukan merupakan bagian dari instrumen pengukuran tetapi lebih melekat pada interpretasi serta penggunaan skor yang dihasilkan oleh suatu

alat ukur pada subjek yang relevan. Perlu adanya terobosan pengembangan pemahaman pada para peneliti atau penyusun alat ukur tentang konsepsi validitas.

Implementasi konsepsi validitas bukanlah persoalan sederhana. Sejak munculnya konsepsi validitas oleh Cronbach dan Meehl (1955), penelitian-penelitian yang memfokuskan pada proses validasi terhadap alat ukur di bidang pendidikan dan psikologi terus berkembang.

Proses validasi yang banyak dilakukan oleh peneliti adalah pengujian validitas isi dan validitas konstruk. Pengujian validitas isi dengan menggunakan pendekatan Lawshe's CVR (*Content Validity Ratio*) atau Aiken's V, sedangkan pengujian validitas konstruk dapat dilakukan dengan pendekatan multitrait-multimethod atau pendekatan analisis faktor dengan menggunakan metode *Explanatory Factor Analysis (EFA)* dan *Confirmatory Factor Analysis (CFA)*. Namun hal yang lebih penting bukanlah proses pengujian validitas tetapi lebih kepada interpretasi skor hasil validitas.

Pada akhirnya 40 tahun kemudian muncul pendapat Messick (1995) yang mendefinisikan validitas sebagai satu kesatuan, tidak terbagi-bagi sebagaimana menurut Cronbach dan Meehl. Messick (1995) menulis bahwa validitas adalah ringkasan evaluatif baik dalam bentuk bukti atau konsekuensi interpretasi dan penggunaan skor hasil tes. Pada saat ini yang lebih diutamakan adalah interpretasi skor hasil tes berdasarkan hasil pengujian validitas.

B. Kajian Pustaka

1. Pengertian Validitas

Validitas berasal dari bahasa Inggris dari kata *validity* yang berarti keabsahan atau kebenaran. Dalam konteks alat ukur atau instrumen asesmen, validitas berarti sejauh mana kecermatan atau ketepatan alat ukur dalam melakukan fungsi ukurnya. Sebuah instrumen yang valid akan menghasilkan data yang tepat seperti yang diinginkan.

Hal tersebut diatas sesuai dengan pendapat Azwar (2013) yang menyatakan bahwa validitas berasal dari kata *validity* yang mempunyai arti sejauhmana akurasi suatu tes atau skala dalam menjalankan fungsi pengukurannya. Pengukuran dikatakan mempunyai validitas yang tinggi apabila menghasilkan data yang secara akurat memberikan gambaran mengenai variable yang diukur seperti yang dikehendaki oleh tujuan pengukuran tersebut, sedangkan pengukuran yang memiliki validitas rendah adalah apabila suatu tes menghasilkan data yang tidak relevan dengan tujuan pengukuran.

Validitas sangat berkaitan dengan tujuan pengukuran. Validitas tidak berlaku secara umum bagi semua pengukuran. Suatu tes mempunyai hasil ukuran yang valid untuk suatu tujuan tertentu yang spesifik tetapi tidak valid untuk tujuan yang lain atau bahkan untuk tujuan yang sama pada kelompok yang lain. Untuk mengetahui alat ukur apakah *valid* atau tidak maka perlu dipelajari dan diukur terlebih dahulu.

Selanjutnya Azwar (2013) juga menjelaskan bahwa validitas adalah pertimbangan yang paling utama dalam mengevaluasi kualitas tes sebagai instrument ukur. Konsep validitas mengacu kepada kelayakan, kebermaknaan dan kebermanfaatan inferensi tertentu yang dapat dibuat berdasarkan skor hasil tes yang bersangkutan. Validasi merupakan proses pengumpulan bukti-bukti yang dapat mendukung inferensi yang dimaksud, dengan kata lain validasi bukan untuk memvalidasi tes melainkan memvalidasi inferensi mengenai penggunaan hasil ukur secara spesifik.

Pada awalnya konsep validitas dianggap sebagai suatu karakteristik yang dimiliki oleh tes, dan saat ini konsep validitas telah mengalami pergeseran dari makna validitas tes menjadi validitas interpretasi skor tes. Validitas lebih diartikan sebagai suatu karakteristik interpretasi skor tes, bukan karakteristik tes ataupun karakteristik skor tes.

Hal tersebut sesuai dengan pengertian validitas yang disampaikan dalam dokumen AERA, APA & NCME (1999) tentang artikel *Standards for educational and psychological tests and manuals* yang diterbitkan oleh *American Psychological Association (APA)*, *American Educational Research Association (AERA)*, dan *National Council on Measurement in Education (NCME)* yang merumuskan konsep validitas adalah sebagai berikut:

- a. Validitas bukanlah karakteristik atau kualitas yang melekat pada tes melainkan kualitas konsekuensi sosial yang ditimbulkan oleh penafsiran hasil tes sesuai tujuan penggunaan tes. Dengan kata lain, validitas dipahami sebagai taraf sejauhmana bukti bukti empiris maupun teoritis mendukung atau membenarkan cara menginterpretasi skor tes sesuai dengan tujuan penggunaan tes. Pengujian validitas yang dilakukan adalah untuk mengevaluasi kualitas interpretasi skor tes sesuai dengan tujuan penggunaan tes, bukan tesnya sendiri.
- b. Validitas sebagai konsep tunggal. Beberapa bukti yang digunakan untuk mengevaluasi kualitas interpretasi skor tes sesuai tujuan penggunaan tes memang mampu menunjukkan aspek-aspek validitas namun tidak mewakili jenis-jenis validitas yang berbeda. Hal ini mengandung makna bahwa validitas merupakan konsep tunggal, yaitu taraf sejauhmana seluruh bukti yang berhasil dikumpulkan mendukung interpretasi skor tes sesuai yang dimaksudkan oleh tujuan penggunaan tes.
- c. Terdapat 5 (lima) jenis bukti yang perlu dikumpulkan untuk memeriksa validitas interpretasi skor sesuai tujuan penggunaan tes yaitu (1) bukti terkait isi tes, (2) bukti terkait proses respon yang diberikan subjek, (3) bukti terkait struktur internal tes, (4) bukti terkait hubungannya dengan variabel lain, dan (5) bukti terkait konsekuensi pengetesan.

Selanjutnya, Linn & Gronlund (2000) mengemukakan hakikat validitas adalah sebagai berikut:

- a. Validitas menyatakan ketepatan interpretasi hasil bukan pada prosedurnya.
- b. Validitas merupakan persoalan yang berkaitan dengan derajat (tingkatan), sebagai konsekuensinya kita harus menghindari pemikiran hasil asesmen sebagai valid atau tidak valid. Oleh karena validitas adalah persoalan derajat maka sebuah instrumen dapat dikategorikan mempunyai derajat validitas tinggi, sedang, dan rendah.
- c. Validitas selalu bersifat khusus untuk penggunaan atau interpretasi tertentu. Tidak ada asesmen yang valid untuk semua tujuan. Sebagai contoh, hasil tes aritmatika mungkin mempunyai tingkat validitas yang tinggi untuk kemampuan hitung, validitas yang rendah untuk alasan-alasan aritmatika, dan mempunyai derajat validitas sedang untuk memprediksi kesuksesan prestasi matematika yang akan datang.
- d. Validitas merupakan kesatuan konsep. Hakikat konsep validitas dipandang sebagai sebuah kesatuan konsep berdasarkan berbagai macam bagian dari fakta.
- e. Validitas melibatkan sebuah keputusan evaluatif yang menyeluruh.

Perubahan pemahaman konsep validitas dari pemahaman awal bahwa validitas itu melekat pada tes bergeser menjadi validitas interpretasi skor tes tentu saja membawa perubahan cara pandang terdapat validitas. Hal ini merupakan proses pengembangan konsepsi terhadap validitas itu sendiri, dan perubahan konsepsi ini juga mempengaruhi proses pengumpulan bukti-bukti empiris pada proses validasi suatu alat ukur atau suatu tes. Tentu saja perubahan konsepsi validitas ini menjadi langkah awal pengembangan dalam pengujian atau pemeriksaan validitas alat ukur atau alat tes.

2. Metode Estimasi Validitas

Dari metode estimasi yang disesuaikan dengan sifat dan fungsi setiap tes, tipe validitas secara tradisional dapat dibedakan menjadi 3 macam yaitu validitas isi (*content validity*), validitas konstruk (*construct*

validity), dan validitas berdasar kriteria (*criterion related validity*) yang terbagi menjadi dua macam yaitu, validitas konkuren (*concurent validity*) dan validitas prediktif (*predictive validity*). Azwar (2013) memberikan penjelasan mengenai 3 (tiga) validitas adalah sebagai berikut:

a. Validitas Isi (*Content Validity*)

- (1) Validitas isi erat kaitannya dengan materi yang akan diukur dalam tes. Tentu saja materi yang dimaksud adalah materi yang terdapat dalam kurikulum. Validitas isi mencerminkan sejauh mana butir-butir dalam tes mencerminkan materi yang disajikan dalam kurikulum. Sebuah tes dikatakan memiliki validitas isi jika butir-butir tes bersifat representatif terhadap isi materi dalam kurikulum tersebut.
- (2) Cara yang praktis untuk melakukan analisis validitas isi adalah dengan melihat apakah butir-butir tes telah disusun sesuai dengan *blue-print* (kisi-kisi) yang sudah dirancang sebelumnya. *Blue print* menjadi acuan dalam menuangkan domain atau ranah dan indikator yang akan diukur dalam tes.
- (3) Pengujian validitas isi secara statistik dapat dilakukan dengan menggunakan formula yang diusulkan oleh Aiken (1985) untuk menghitung koefisien validitas isi yang didasarkan pada hasil penilaian dari panel ahli sebanyak n orang terhadap suatu aitem dari segi sejauhmana aitem tersebut mewakili konstruk yang diukur. Penilaian dilakukan dengan cara memberikan angka antara 1 (sangat tidak mewakili/sangat tidak relevan) sampai dengan 5 (sangat mewakili/sangat relevan). Bila terdapat sebanyak n panelis yang menilai sebuah aitem melalui rating (r) dengan pilihan 1 (sangat tidak relevan) sampai dengan sangat relevan (5), berarti kategori tertinggi (c) adalah 5 dan kategori terendah (l) adalah 1, maka dapat

dituliskan dalam rumus sebagai berikut:

$$V = \sum s / [n(c - l)] \quad (1)$$

Keterangan:

$$s = r - l$$

V memiliki kemungkinan nilai 0 s.d. 1 yang menunjukkan derajat validitas aitem. Sebuah aitem dianggap valid manakala memiliki V sebesar 0.5 atau lebih.

- (4) Pengujian validitas isi secara statistik yang kedua adalah menurut Lawshe (1975) yang dikenal dengan rasio validitas isi atau content validity ratio (CVR). Pendekatan yang dilakukan adalah dengan melibatkan subject matter expert (SME) diminta untuk menyatakan apakah aitem dalam tes sifatnya esensial bagi operasional konstruk teoritik tes yang bersangkutan. Formula persamaannya adalah:

$$CVR = (2 ne / n) - 1 \quad (2)$$

Keterangan:

Ne : banyaknya SME yang menilai suatu aitem esensial

N : banyaknya SME yang melakukan penilaian.

Perhitungan formula tersebut akan menghasilkan skor CVR yang terentang dari -1 s.d. 1. Bila setengah dari panelis menyatakan sebuah aitem bersifat esensial, $CVR = 0$, berarti aitem tersebut valid.

b. Validitas Konstrak (*Construct Validity*)

- (1) Validitas konstrak adalah validitas yang menyangkut bangunan

- teoretik variabel yang akan diukur. Sebuah tes dikatakan mempunyai validitas konstrak apabila butir-butir soal yang disusun dalam tes mengukur setiap aspek berpikir dari sebuah variabel yang akan diukur melalui tes tersebut.
- (2) Seperti halnya validitas isi, untuk mempertinggi validitas konstrak dapat dilakukan dengan cara merinci dan memasang setiap butir soal dengan setiap aspek. Pengujian validitas konstrak diperlukan analisis statistik yang kompleks seperti prosedur analisis faktor. Dalam analisis faktor dikenal dua macam prosedur yang dilandasi oleh dasar pemikiran yang berbeda yaitu *exploratory factor analysis* (EFA) yang akan membantu penyusun tes untuk mengenali dan mengidentifikasi bebrbagai faktor yang membentuk suatu konstrak dengan cara menemukan varians skor terbesar dengan jumlah faktor yang paling sedikit yang dinyatakan dalam bentuk $eigenvalue > 0,1$ dan *confirmatory factor analysis* (CFA) yang biasanya dilakukan sebagai tindaklanjut dari hasil EFA dengan menyertakan dasar teori yang melandasi bangunan tes yan bersangkutan agar lebih lanjut dapat menguji validitas konstraknya.
 - (3) Salah satu prosedur pengujian validitas konstrak yang tidak terlalu kompleks dapat dilakukan dengan pendekatan *multitrait-multimethod*. Dua atau lebih *trait* yang diukur melalui dua atau lebih metode dapat diuji secara serentak dengan pendekatan ini, sehingga akan diperoleh adanya bukti adanya validitas diskriminan dan validitas konvergen.
 - (4) Validitas diskriminan ditunjukkan oleh rendahnya korelasi antara faktor skala atau tes yang mengukur *trait* yang berbeda terutama bila digunakan metode yang sama. Validitas konvergen ditunjukkan oleh tingginya korelasi skor tes-tes yang mengukur *trait* yang sama dengan menggunakan metode yang berbeda.
- c. Validitas Berdasarkan Kriteria
- (1) Validitas kriteria ini didasarkan pada kriteria tertentu. Dengan demikian bukti adanya validitas ditunjukkan adanya hubungan korelasional skor pada tes yang bersangkutan dengan skor suatu kriteria.
 - (2) Pengujian validitas ini bersifat empirik, artinya pengujian hanya dapat dilakukan setelah mendapatkan data di lapangan. Apabila berdasarkan hasil analisis yang dilakukan terhadap data hasil pengamatan di lapangan terbukti bahwa tes hasil belajar dapat mengukur hasil belajar yang seharusnya diungkap secara tepat maka berarti alat tes tersebut mempunyai validitas empirik. Untuk keperluan pengujian validitas kriteria dapat dilakukan dengan dua cara yaitu dari segi ketepatan meramalkan (*predictive validity*) dan ketepatan bandingannya (*concurrent validity*).
 - (3) Perbedaan utama antara validitas ramalan dengan validitas bandingan adalah ketersediaan pembanding (kriterium). Pada validitas ramalan, kriterium diperoleh pada waktu yang akan datang setelah dilakukan tes yang akan diukur validitasnya tersebut. Sedangkan pada validitas bandingan, kriterium sudah ada atau dapat diperoleh pada saat yang sama dengan waktu untuk memperoleh data tentang tes yang akan diukur validitasnya tersebut tanpa harus menunggu masa yang akan datang.
 - (4) Validitas ramalan (*predictive validity*). Sebagai ilustrasi adalah adanya tes masuk Perguruan Tinggi Negeri. Setelah melalui serangkaian tes maka hanya calon mahasiswa yang mempunyai skor tinggi yang

diterima oleh panitia seleksi mahasiswa baru. Sesungguhnya keputusan panitia seleksi yang hanya menerima mahasiswa yang mempunyai skor tinggi saja berarti sudah memprediksikan bahwa calon mahasiswa dengan skor tinggi tersebut kelak yang akan lebih berhasil dalam studinya. Sedangkan para calon mahasiswa yang mempunyai skor sedang apalagi rendah diprediksikan akan banyak menemui kendala dalam studinya. Oleh karena itu tes yang digunakan dalam seleksi calon mahasiswa baru tersebut akan mempunyai tingkat validitas prediktif yang tinggi apabila secara empirik terbukti bahwa prestasi belajar mereka juga baik. Dengan demikian antara skor tes masuk dengan prestasi belajar harus mempunyai korelasi yang positif. Pada kasus tersebut, yang dipermasalahkan validitasnya adalah tes masuk. Oleh karena itu hasil belajar pada masa perkuliahan digunakan sebagai tolok ukur (kriterium). Adanya kesejajaran, kesesuaian, kesamaan arah antara tes seleksi masuk dengan hasil belajar mempunyai korelasi yang positif.

- (5) Validitas bandingan (*Concurrent Validity*). Validitas ini sering pula disebut sebagai validitas ada sekarang (konkuren), validitas sama saat, validitas pengalaman, dan validitas empiris. Disebut sebagai validitas ada sekarang karena pengujiannya berdasarkan pengalaman yang saat ini sudah ada di tangan. Disebut sebagai validitas sama saat karena validitas ini segera dapat kita peroleh informasinya pada saat yang sama dengan waktu diperolehnya data hasil tes yang diukur validitasnya tersebut. Disebut validitas pengalaman (empiris) karena validitas ini dikaitkan dengan pengalaman yang sudah ada. Dalam

hal ini pengalaman digunakan sebagai kriterium. Guna menentukan validitas bandingan ini tidak perlu menunggu waktu untuk membuktikannya. Dalam validitas bandingan ini menunjukkan bahwa yang berfungsi sebagai kriterium adalah data hasil pengalaman. Apabila data dari tes yang ada sekarang mempunyai hubungan yang searah dengan data hasil pengalaman maka dikatakan telah mempunyai validitas bandingan.

Pemahaman yang baru tentang validitas beserta metode estimasinya, khususnya setelah terbitnya artikel AERA, APA & NCME tahun 1999 yang lebih menekankan sejauhmana interpretasi skor tes sebagaimana dimaksud oleh tes yang bersangkutan sungguh-sungguh dapat dipertanggungjawabkan. Berdasarkan konsep validitas sebagai interpretasi skor tes maka diperlukan 5 (lima) jenis bukti yang dikumpulkan dalam memeriksa validitas interpretasi skor tes atau hasil pengukuran suatu tes sesuai dengan tujuan penyusunan tes tersebut.

AERA, APA & NCME (1999) dan Goodwin & Leech (2003) menjelaskan kelima bukti yang perlu dikumpulkan adalah sebagai berikut:

- a. Bukti terkait isi tes
Bukti tentang kesesuaian isi dan konstruk yang diukur oleh suatu tes dapat diperoleh melalui analisis logis atau empiris terhadap seberapa memadai isi tes mewakili ranah isi serta seberapa relevan ranah isi tersebut sesuai dengan interpretasi skor tes yang dimaksudkan. Isi tes mengacu pada tema-tema, pilihan kata, serta format atau bentuk aitem, tugas atau pertanyaan yang digunakan dalam tes.

Bukti terkait isi diperoleh melalui penilaian pakar atau ahli terhadap kesesuaian anatara bagian-bagian tes dan konstruk yang diukur. Aspek-aspek isi tes yang perlu dievaluasi meliputi;

- (1) *Sufficiency* atau kecukupan yaitu

apakah isi tes tersebut mencukupi atau memadai/mewakili ranah isi spesifik yang hendak diukur.

- (2) *Clarity* atau kejelasan yaitu apakah isi tes tersebut mencerminkan secara jelas ranah isi spesifik yang hendak diukur, misalnya tidak mencampuradukkan dengan ranah isi spesifik yang lain.
- (3) *Relevance* atau relevansi yaitu apakah isi tes tersebut memiliki kesesuaian dengan ranah isi spesifik yang hendak diukur.
- (4) Kesesuaian antara aitem-aitem dan tugas-tugas yang dipakai sebagai stimulus dalam tes tersebut dengan definisi tentang konstruk yang hendak diukur.
- (5) Ada tidaknya bias berupa keberpihakan isi tes pada gender, budaya, umur atau faktor pengelompokan sosial lainnya.
- (6) Kemungkinan terjadinya varians yang tidak relevan dengan konstruk yang hendak diukur

Hal-hal yang perlu dilakukan untuk memperhatikan validitas yang terkait dengan isi adalah menyusun kisi-kisi atau tabel spesifikasi, langkah kedua yaitu melakukan eksplikasi konstruk dengan cara perumusan definisi konseptual konstruk dan merumuskan definisi operasional konstruk, dan langkah ketiga adalah melakukan analisis tugas, jika terkait dengan pengukuran yang kompleks dan diperlukan adanya *job analysis* untuk membantu membuat komponen dan uraian perilaku yang lebih spesifik sesuai dengan komponen tugas.

- b. **Bukti terkait proses respon subjek**
Bukti ini didasarkan pada penilaian terhadap kesesuaian antara respon yang diberikan oleh subjek dalam rangka mengerjakan tes dengan konstruk yang diukur oleh tes. Strategi untuk mengumpulkan bukti ini adalah melakukan observasi terhadap subjek saat tes berlangsung atau mewawancarai subjek untuk mengetahui alasan memberikan jawaban tertentu terhadap pertanyaan-

pertanyaan dalam tes.

- c. **Bukti terkait struktur internal tes**
Bukti ini didasarkan pada penilaian sejauhmana aitem-aitem dan komponen-komponen dalam tes saling berhubungan sedemikian rupa sesuai dengan konstruk yang diukur. Hal ini terkait dengan konsistensi internal atau homogenitas tes. Langkah yang dapat dilakukan untuk memeriksa struktur internal tes adalah analisis faktor konfirmatori dan *differential item function* (DIF) untuk memeriksa kemungkinan terjadinya bias aitem.
- d. **Bukti terkait hubungan antara tes dengan tes lain**
Bukti ini didapatkan dengan menganalisis hubungan antara skor tes dan variabel-variabel lain di luar tes itu sendiri. Metode pertama yang dapat dilakukan adalah analisis hubungan antara skor tes dan skor kriteria yang diprediksikan oleh tes yang bersangkutan, yang kedua adalah melakukan analisis hubungan antara skor tes dan skor tes-tes lain yang dimaksudkan untuk mengukur konstruk yang sama seperti yang diukur oleh tes yang bersangkutan dan analisis hubungan antara skor tes dan skor tes-tes lain yang dimaksudkan untuk mengukur konstruk yang berbeda dari yang diukur oleh tes yang bersangkutan. Langkah ini akan memberikan bukti konvergen dan diskriminan. Langkah ketiga adalah analisis perbedaan kinerja dalam tes yang sama antara dua atau lebih kelompok subjek yang diprediksikan memang akan berbeda berkat hubungan antara konstruk yang diukur oleh tes dan variabel yang mendasari pembagian subjek dalam kelompok-kelompok.
- e. **Bukti terkait konsekuensi pengetesan**
Bukti ini terkait dengan konsekuensi, dampak atau akibat dari proses administrasi tes terhadap kinerja atau perilaku subjek. Konsekuensi pengetesan ini dibedakan dalam dua

kategori yaitu konsekuensi yang direncanakan dan konsekuensi yang tidak direncanakan.

3. Faktor-faktor yang mempengaruhi validitas

Faktor-faktor yang mempengaruhi validitas suatu alat ukur atau alat tes dapat bersumber pada alat tes itu sendiri dan faktor dari luar. Faktor-faktor yang mempengaruhi validitas antara lain:

a. Faktor dari dalam tes itu sendiri

Pengujian terhadap aitem dalam tes secara hati-hati akan menunjukkan apakah tes yang digunakan untuk mengukur isi materi atau fungsi-fungsi mental yang akan diukur oleh penyusun tes. Bagaimanapun juga, beberapa faktor berikut dapat menjaga aitem tes dari fungsi yang dikehendaki dan dengan demikian juga terjaga dari rendahnya validitas. Berikut hal-hal yang terkait dengan faktor yang bersumber dari dalam tes itu sendiri:

- (1) Petunjuk yang tidak jelas. Petunjuk yang tidak jelas menyebabkan subjek kehilangan waktu untuk sekedar memahami petunjuk pengerjaan atau bahkan tidak dapat melakukan apa yang seharusnya dilakukan.
- (2) Penggunaan kosa kata dan struktur kalimat yang sulit. Penggunaan kosa kata atau struktur kalimat yang sulit dapat menyebabkan subjek terjebak untuk pemahaman terhadap pemahaman maksud dari sebuah pertanyaan bukan untuk menyelesaikan pertanyaan itu sendiri.
- (3) Ambiguitas yaitu adanya kemungkinan multi tafsir juga menyebabkan menurunnya validitas.
- (4) Alokasi waktu yang tidak cukup. Idealnya sebuah tes disediakan waktu yang cukup untuk mengerjakan seluruh butir tes yang ada. Kekurangan waktu dalam menyelesaikan sebuah tes bisa jadi bukan karena subjek tidak mampu untuk menyelesaikan tesnya tetapi karena keterbatasan kesempatan untuk mengerjakannya.

- (5) Penekanan yang berlebihan terhadap aspek tertentu, sehingga terlalu mudah ditebak kecenderungan dari jawaban akan menyebabkan menurunnya tingkat validitas soal.
- (6) Kualitas butir tes yang tidak memadai untuk mengukur hasil belajar. Kualitas yang tidak memadai misalnya tes dimaksudkan untuk mengukur kemampuan berpikir tingkat tinggi (*higher order thinking*) jelas tidak cukup hanya digunakan tes yang bersifat untuk mengungkap pengetahuan faktual saja.
- (7) Susunan tes yang jelek.
- (8) Tes terlalu pendek.
- (9) Penyusunan butir tes yang tidak runtut.
- (10) Pola jawaban yang mudah ditebak, misalnya pada soal pilihan ganda jawabannya adalah A semua, atau B semua atau menunjukkan pola tertentu misalnya D, C, B, A, D, C, B, A, dan sebagainya.

b. Faktor administrasi dan penskoran.

Pemberian skor terhadap jawaban subjek (*testee*) harus dilakukan secara hati-hati jangan sampai salah tulis atau meremehkan selisih angka walaupun hanya sedikit. Hal ini akan menyebabkan hasil pengujian terhadap validitas akan memberikan makna yang berbeda.

c. Faktor tanggapan subjek.

Tanggapan subjek yang tidak serius biasanya dijumpai pada saat subjek diminta untuk mengisi sebuah angket atau skala. Hal ini akan menyebabkan subjek mengisi angket atau skala secara sembarangan karena merasa tidak penting maupun alasan-alasan yang lain. Oleh karena itu berikan angket/skala pada waktu dan kondisi yang tepat.

d. Hakikat kelompok dan kriteria.

Seperti sudah dijelaskan di atas bahwa validitas bersifat spesifik. Sebuah asesmen atau instrumen alat ukur mungkin hanya valid untuk kelompok tertentu saja dan tidak valid untuk kelompok yang lain. Sebagai contoh misalnya sebuah tes diujicobakan pada sekelompok subjek pada sebuah sekolah dengan kualitas biasa-biasa saja tentu akan berbeda hasilnya jika tes yang sama diberikan pada sekelompok

subjek pada sekolah yang favorit.

C. Implementasi Pengujian Validitas Isi dan Validitas Kontrak dalam Konteks Validitas Interpretasi Skor.

Dalam naskah ini, contoh implementasi pengujian validitas isi dan kontrak dalam konteks sebagai validitas interpretasi skor menggunakan instrumen Indonesia Spiritual Health Assessment (ISHA) yang dikembangkan oleh Centre for Neuroscience, Health, and Spirituality (C-NET) UIN Sunan Kalijaga Yogyakarta. ISHA disusun berdasarkan 5 dimensi yaitu (1) pengalaman spiritual, (2) emosi positif, (3) makna hidup, (4) ritual dan (5) dominansi otak.

Penelitian yang dilakukan melibatkan 1000 responden yang terdiri dari 10 kota besar yang ada di Indonesia yaitu Jakarta, Bandung, Surabaya, Semarang, Yogyakarta, Manado, Manokwari, Banjarmasin, Pontianak, dan Palembang. Analisis statistik yang dilakukan adalah sebagai berikut;

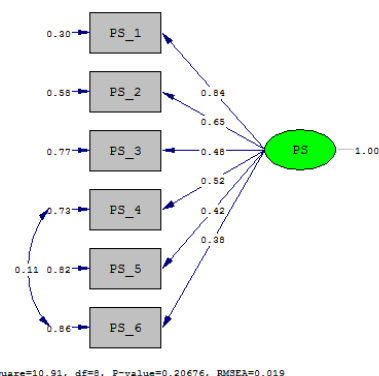
1. Exploratory Factor Analysis dan Confirmatory Factor Analysis menggunakan SPSS dan Lisrel dengan tujuan untuk memperoleh informasi tentang uni-dimensionality instrument, kecocokan (fitness) dari model pengukuran dan rekomendasi model pengukuran yang fit.
2. Analisis butir aitem dengan pendekatan teori tes klasik dengan menggunakan Iteman, untuk memperoleh karakteristik setiap aitem, statistic tes untuk setiap dimensi, konsistensi internal (alpha Cronbach) setiap dimensi dan rekomendasi dalam seleksi aitem.
3. Analisis butir dengan pendekatan model RASCH dengan software Winsteps. Analisis dilakukan untuk memperoleh data distribusi aitem-responden dalam skala yang sama, kecocokan aitem dan responden terhadap model (fitness), skala baku setiap aitem dan rekomendasi dalam

seleksi ke dalam bank soal.

Hasil dan kesimpulan analisis terhadap instrument ISHA adalah sebagai berikut;

1. Dimensi Pengalaman Spiritual

- a. Berdasarkan kriteria unidimensi dengan menggunakan eigen value: (1) Eig-1 dibagi total eig > 20% dan (2) Eig-1 dibagi eig 2 >4. Terlihat bahwa 60 item pertama tidak dapat dikatakan sebagai unidimensi. Satu *common factor* utama hanya mampu menjelaskan 11.18% dari total variance. Di samping itu eig-1 dibagi dengan eig-2 nilainya kurang dari 4.
- b. Jika disesuaikan dengan desain yang memiliki 6 subdimensi, maka dari EFA terlihat bahwa variance yang mampu dijelaskan ada 25.59% Berdasarkan hasil EFA 6 subdimensi pengalaman spiritual, 60 aitem menjadi 59 (aitem nomoe 25 drop)
- c. Berdasarkan distribusi aitem pada *rotated component matrix*, maka dilakukan pembentukan variabel baru berupa sub dimensi pengalaman spiritual 1 sampai 6. Setiap sub dimensi merupakan jumlahan dari response aitem-aitem yang membentuk sub dimensi tersebut.
- d. Selanjutnya dilakukan EFA pada 6 variabel sub dimensi, dengan hasil tampak bahwa komponen pertama mampu menjelaskan 42% variance dengan hasil analisis CFA adalah sebagai berikut:

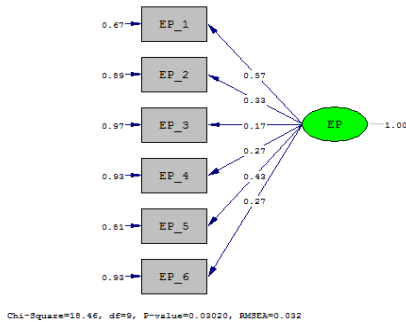


- e. Dari hasil analisis CFA di atas tampak bahwa model fit unidimensi diperoleh jika pengelompokan item pada setiap

subdimensi mengikuti hasil EFA, dengan melakukan korelasi antara error subdimensi PS_4 dengan PS_6.

2. Dimensi Emosi Positif

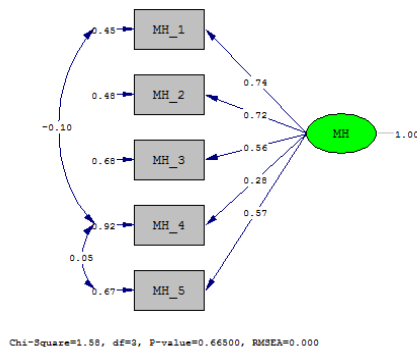
a. Hasil EFA dan CFA dimensi Emosi Positif adalah sebagai berikut:



b. Berdasarkan hasil CFA tampak bahwa respons terhadap 60 item yang menyusun dimensi emosi positif tidak cocok dengan model satu dimensi, meskipun setiap subdimensi memiliki factor loading yang baik.

3. Dimensi Makna Hidup

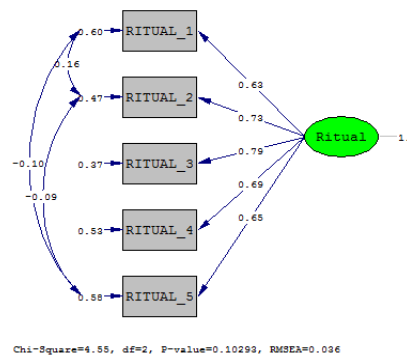
a. Hasil EFA dan CFA dengan menggunakan 5 faktor:



b. Kesimpulan: Dimensi makna hidup akan bersifat unidimensi jika error variance dari MH-1 dan MH-4 dikorelasikan begitu juga dengan error variance MH-4 dan MH-5

4. Dimensi Ritual

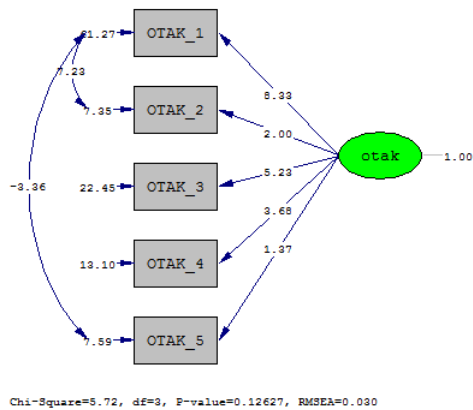
- a. Menyusun 5 variabel baru yang merupakan komposit dari soal-soal yang menyusun subdimensi ritual yaitu: (1) melakukan sembahyang, (2) berdoa, (3) mengunjungi tempat ibadah, (4) bersedekah, dan (5) terlibat dalam aktivitas/komunitas keagamaan. |
- b. Dari hasil EFA tampak bahwa kelima subdimensi tersebut bersifat unidimensi. Selanjutnya analisis CFA dengan memberikan hasil sebagai berikut:



c. Hasil CFA menunjukkan bahwa data FIT bersifat unidimensi jika error subdimensi melakukan RITUAL_1 (sembahyang) berkorelasi dengan error sub dimensi RITUAL_2 (berdoa) dan subdimensi RITUAL_5 (terlibat dalam komunitas). Selain itu error sub dimensi RITUAL_2 (berdoa) juga dikorelasikan dengan error sub dimensi RITUAL_5 (terlibat dalam komunitas).

4. Dominansi Otak

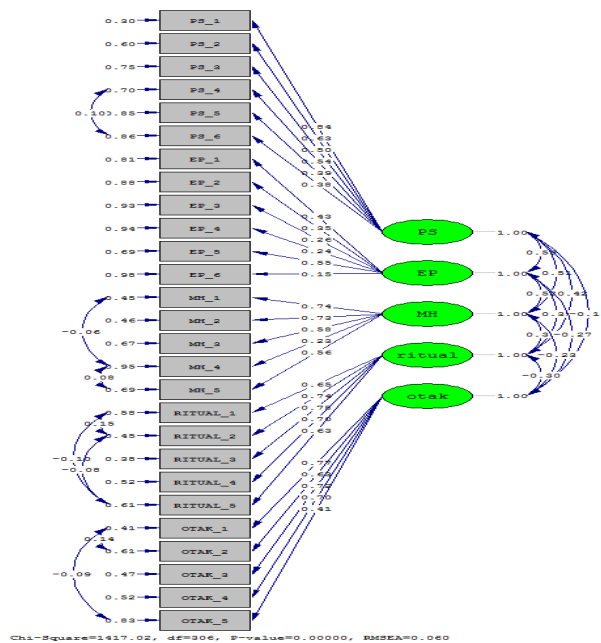
- a. Dari analisis EFA diperoleh component matrix yang menunjukkan dimensi dominansi otak terdiri dari 5 komponen. Berdasarkan pengelompokan item pada EFA, disusun 5 variabel komposit baru.
- b. Hasil analisis menggunakan CFA adalah sebagai berikut:



dengan error dimensi OTAK_2, serta error dimensi OTAK_1 dikorelasikan dengan error dimensi OTAK_5.

5. Hasil *second order* CFA menunjukkan 5 dimensi ISHA yaitu pengalaman spiritual, emosi positif, makna hidup, ritual dan dominansi otak tidak bersifat unidimensi, sehingga disarankan hasil tes ISHA dilaporkan dalam 5 sub skor yang mewakili masing-masing dimensi Pengalaman Spiritual, Emosi Positif, Makna Hidup, Ritual dan Dominansi Otak.

- c. Kesimpulannya adalah model pengukuran akan FIT unidimensi jika error dimensi OTAK_1 dikorelasikan



6. Hasil analisis aitem dengan menggunakan iteman adalah sebagai berikut:

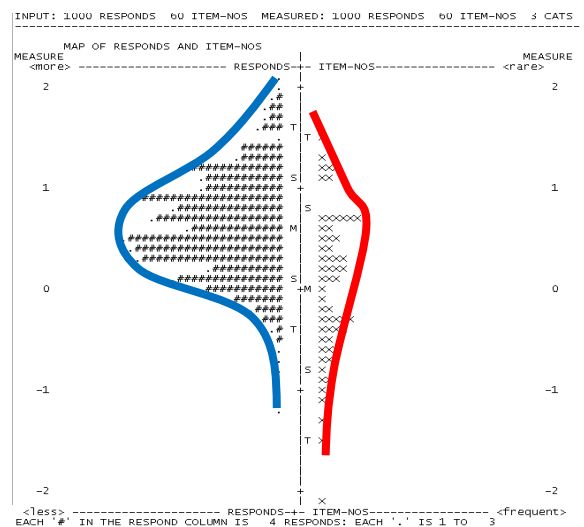
PROCEEDING
Seminar Nasional Psikometri

REKAPITULASI HASIL ANALISIS ITEMAN						
DIMENSI	JML. AITEM	DITOLAK	DITERIMA	NOMOR AITEM	KETERANGAN	
A-1	SPIRITUAL	60	4	56	1, 3, 21 dan 25	Check the key *)
A-2	EMOSI POSITIF	60	4	56	8, 31, 47, dan 50	Daya beda < 0.20
A-3	MAKNA HIDUP	60	1	59	5	Daya beda < 0.20
A-4	RITUAL	60	16	44	5, 6, 12, 14, 17, 19, 21, 22, 23	Daya beda < 0.20
					27, 28, 34, 40, 44, 52, 56	Daya beda < 0.20
		240	25	215	22	Check the key *)
DIMENSI	JML. AITEM	DITOLAK	DITERIMA	NOMOR AITEM	KETERANGAN	
B-1	CORTEX PREFONTALIS	15	1	14	13	Check the key *)
B-2	SISTEM LIMBIK	15	0	15		
B-3	GANGLIA BASALIS	12	0	12		
B-4	GYRUS CINGULATUS	9	0	9		
B-5	LOBUS TEMPORALIS	12	0	12		
		63	1	62		

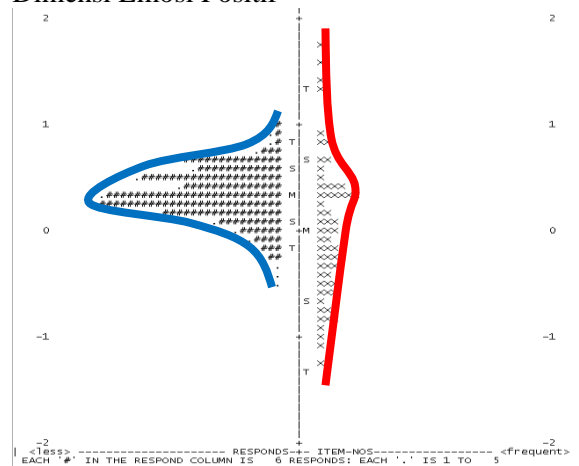
*) Berdasarkan nilai ITEM-SCALE CORRELATION, ITEMAN mencurigai adanya permasalahan konten atau ketidak-sesuai-an dalam penetapan favorable/unfavorable

Berdasarkan hasil analisis menggunakan Iteman, dapat disimpulkan bahwa terdapat 4 aitem yang ditolak (gugur) pada dimensi pengalaman spiritual dan terdapat catatan untuk melihat kesesuaian kunci jawaban karena adanya kecurigaan permasalahan isi dari aitem. Pada dimensi emosi positif terdapat 4 aitem yang gugur dan pada dimensi makna hidup gugur 1 aitem, sedangkan pada dimensi ritual terdapat 16 aitem yang gugur dan pada aitem nomor 22 perlu dilakukan pengecekan kesesuaian kunci jawaban dan pada dimensi dominansi otak hanya 1 aitem yang gugur dengan adanya kecurigaan terhadap kesesuaian kunci jawaban. Jumlah aitem yang gugur adalah 26 dan jumlah aitem yang diterima sebanyak 277 dari 303 aitem yang dibuat.

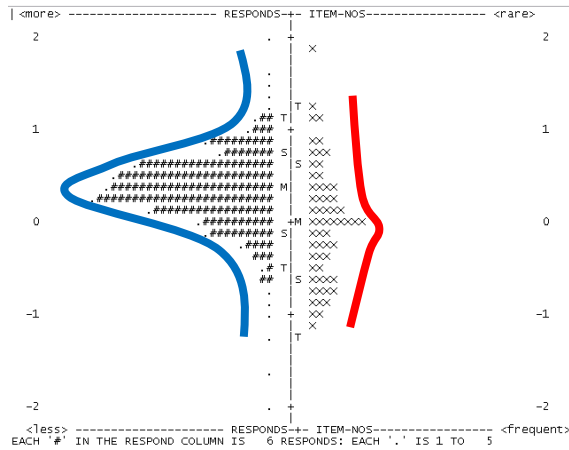
7. Hasil analisis aitem dengan menggunakan model RASCH adalah sebagai berikut:
 - a. Dimensi Pengalaman Spiritual



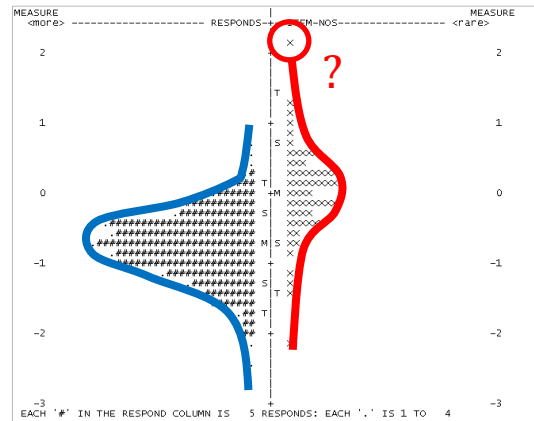
b. Dimensi Emosi Positif



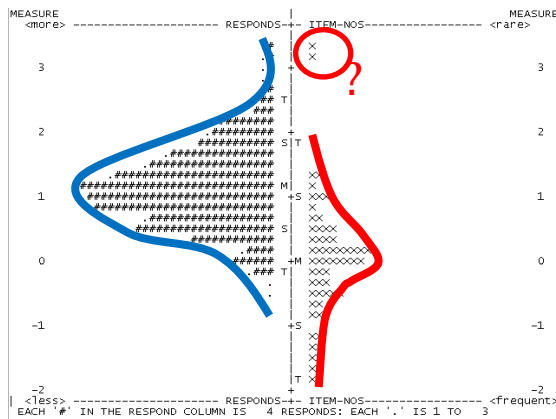
c. Dimensi Makna Hidup



e. Dimensi Dominansi Otak



d. Dimensi Ritual



Hasil analisis aitem dengan menggunakan model RASCH adalah sebagai berikut:

REKAPITULASI HASIL ANALISIS WINSTEPS												
DIMENSI	AITEM	AITEM					KET	RESPONDEN				
		MISFIT	MEAN	S.DEV.	MIN	MAX		MISFIT	MEAN	S.DEV.	MIN	MAX
A-1	SPIRITUAL	0	0	0.78	-2.34	1.54		14	0.63	0.52	-1.4	2.8
A-2	EMOSI POSITIF	0	0	0.66	-1.23	1.75		5	0.35	0.25	-0.6	1.1
A-3	MAKNA HIDUP	0	0	0.63	-1.19	1.92		6	0.34	0.72	-0.8	2
A-4	RITUAL	2	0	1.15	-2.09	4.01	AITEM 7 & 22	93	1	0.72	-0.8	4
B	DOI - 5 DIMENSI	0	0	0.77	-2.43	2.46		54	-0.87	0.51	-3.4	0.4
B-1	CORTEX PREFONTALIS	0	0	0.75	-2.44	0.96		58	-0.76	0.61	-3.4	0.9
B-2	SISTEM LIMBIK	0	0	1.21	-2.01	3.47		96	-0.97	0.77	-4.6	1.2
B-3	GANGLIA BASALIS	0	0	0.8	-1.92	1.22		73	-1.02	0.84	-5	1.2
B-4	GYRUS CINGULATUS	0	0	0.69	-1.34	1.05		95	-0.69	0.82	-5	2.2
B-5	LOBUS TEMPORALIS	0	0	0.73	-1.46	1.17		63	-1.83	1.06	-5	0.08

*) JIKA NILAI MNSQ OUTFIT > 2.0 maka ITEM/PERSON TSB TIDAK FIT (MISFIT) DENGAN MODEL RASCH

Berdasarkan hasil analisis aitem dengan menggunakan pendekatan RASCH, dapat disimpulkan bahwa

terdapat 2 aitem yang misfit pada dimensi ritual yaitu aitem nomor 7 dan 22, sedangkan pada dimensi

pengalaman spiritual, emosi positif, makna hidup dan dimensi dominansi otak semua aitem mempunyai model fit dengan model pengukuran yang disusun. Dari sisi responden diketahui terdapat beberapa responden yang misfit, hal ini menunjukkan adanya bias pada responden.

4. Hasil pengujian validitas konstrak dengan menggunakan model RASCH menunjukkan bahwa ITEM Measure (skala aitem) berlaku sebagai karakteristik aitem yang menjadi dasar dalam estimasi PERSON Measure

Daftar Pustaka

D. Simpulan

Berdasarkan hasil analisis pengujian validitas interpretasi skor tes yang dihasilkan maka simpulan yang dapat diambil terkait dengan penggunaan ISHA sebagai alat ukur yang bertujuan untuk mengukur kesehatan spiritualitas adalah sebagai berikut;

1. Dari sisi validitas isi dan konstrak menunjukkan bahwa isi dan konstrak penyusunan ISHA sesuai dengan konseptualisasi teori yang digunakan dan definisi operasionalnya yaitu dengan model 5 dimensi dan bersifat multidimensi artinya penggunaan hasil tes ISHA tidak dalam bentuk skor total tetapi mengacu pada masing-masing dimensi sehingga fungsi pengukuran ISHA tepat sebagai salah satu alat ukur yang akan menghasilkan profil kesehatan spiritual seseorang.
2. Penyusun alat ukur perlu melakukan *check* terhadap beberapa aitem berdasarkan informasi hasil analisis pemeriksaan analisis aitem.
3. Seleksi aitem berdasarkan output Iteman maupun model RASCH menunjukkan aitem-aitem yang berkualitas dari sisi indek daya beda yang memenuhi syarat untuk dimasukkan ke dalam bank soal

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

Azwar, S (2013). *Reliabilitas dan validitas-edisi 4*. Yogyakarta: Pustaka Pelajar

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

Goodwin, L.D. & Leech, N.L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counselin and Development*, 36, 181-191.

Linn, R.L, Grondlund, N.E. (2000). *Measurement and assessment in teaching*. Eighth edition. New Jersey: Merrill an imprint of Prentice Hall.

Messick, S. J. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.