

Kriteria Instrumen dalam suatu Penelitian

Zaenal Arifin

Program Studi Pendidikan Matematika, Universitas Majalengka

Email: arifin1169@gmail.com

Abstrak— Artikel ini bertujuan untuk melakukan kajian tentang kriteria instrumen yang baik yang diperlukan dalam suatu penelitian. Validitas, reliabilitas, tingkat kesukaran, daya pembeda dan pengecoh (distraktor) merupakan hal yang penting yang harus diperhatikan oleh peneliti agar instrumen yang dibuat menjadi baik, sehingga bisa digunakan dalam dan menghasilkan data yang baik pula. 1) Validitas terbagi menjadi validitas isi dan konstruk. Validitas isi ditentukan oleh nilai V , sedangkan validitas konstruk ditentukan oleh nilai KMO. 2) Reliabilitas instrumen ditentukan oleh nilai *cronbach's alfa*. 3) Tingkat kesukaran butir soal ditentukan oleh nilai koefisien p . 4) Daya pembeda butir soal ditentukan oleh koefisien DB. 5) Distraktor/pengecoh ditentukan oleh persentase pemlilihan dari peserta tes. Jika nilai atau persentase dari kelima hal ini berada pada kategori baik, maka instrumen yang dibuat peneliti memiliki kriteria yang baik.

Kata kunci: daya pembeda, instrumen yang baik, pengecoh, reliabilitas, tingkat kesukaran, validitas.

1. PENDAHULUAN

Bagi seorang peneliti, mempelajari metodologi penelitian, dalam menentukan data tidaklah mudah. Tidak hanya mendapatkan data dan mengolahnya sesuai keinginan sendiri, namun harus mengikuti prosedur yang bisa dilakukan untuk menganalisis data tersebut. Oleh karena itu, agar data yang kita peroleh menjadi data yang valid dan reliabel, diperlukan adanya suatu instrumen atau yang biasa disebut dengan alat ukur yang baik pula.

Menurut Azwar (2011: 2), menyatakan bahwa para ahli psikometri telah menetapkan kriteria bagi alat ukur, seperti instrumen, untuk dinyatakan sebagai alat ukur yang baik. Kriteria tersebut antara lain adalah valid, reliabel, standar, ekonomis dan praktis. Menurut Gronlund (2009: 70) menyatakan bahwa karakteristik utama yang harus dimiliki oleh sebuah alat ukur dapat diklasifikasikan menjadi karakter validitas, reliabilitas dan tingkat kegunaannya. Secara singkat dan jelas, dalam membuat instrumen penelitian pengembangan, Setyosari (2013: 207) dan Depdiknas (2008: 4) menyatakan bahwa validitas dan reliabilitas menjadi dua hal yang harus benar-benar diperhatikan.

Dengan adanya instrumen penelitian, maka kita akan mengetahui sumber data yang akan kita teliti dan jenis datanya, teknik pengumpulan datanya, instrumen pengumpulan datanya, langkah penyusunan instrumen penelitian tersebut serta mengetahui validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan pengecoh/distraktor suatu data dalam penelitian.

Instrumen mempunyai peranan yang sangat penting. Karena dengan adanya instrumen, mutu suatu penelitian dapat diketahui. Jika instrumen yang dibuat, memiliki

kriteria yang baik, maka mutu penelitiannya juga baik, begitupun sebaliknya.

Hal ini mudah dipahami karena instrumen berfungsi mengungkapkan suatu fakta menjadi suatu data, sehingga jika instrumen yang digunakan dalam penelitian mempunyai kualitas yang baik, dalam arti valid dan reliabel serta memiliki tingkat kesukaran, daya pembeda dan distraktor/pengecoh yang baik, maka data yang diperoleh akan sesuai dengan fakta atau keadaan sesungguhnya di lapangan. Sedangkan jika kualitas instrumen yang digunakan tidak baik dalam arti mempunyai validitas dan reliabilitas yang rendah, serta memiliki tingkat kesukaran, daya pembeda dan distraktor/pengecoh yang tidak baik, maka data yang diperoleh juga tidak valid atau tidak sesuai dengan fakta di lapangan, sehingga dapat menghasilkan kesimpulan yang keliru.

Sebenarnya, dalam mengumpulkan data dalam suatu penelitian, selain membuat instrumen sendiri, kita dapat menggunakan instrumen yang telah tersedia (instrumen baku). Instrumen yang telah tersedia pada umumnya sudah dianggap baku untuk mengumpulkan data terhadap variabel-variabel tertentu.

Oleh karena itu, jika instrumen telah tersedia untuk mengumpulkan data variabel suatu penelitian, maka kita dapat langsung menggunakan instrumen tersebut. Namun ada catatan, bahwa teori yang dijadikan landasan penyusunan instrumen tersebut harus sesuai dengan teori yang dirujuk pada penelitian kita.

Selain itu konstruk variabel yang diukur oleh instrumen baku, harus juga sama dengan konstruk variabel yang akan kita ukur dalam penelitian kita. Akan tetapi, lebih baik jika kita sendiri

yang membuat instrumen. Selain belajar membuat instrumen sendiri, kita juga bisa mengasah kemampuan yang kita punya.

Kajian ini akan membahas mengenai beberapa hal yang harus diperhatikan, agar instrumen yang dibuat bisa dilihat kriterianya, apakah instrumennya baik atau tidak baik, yaitu validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan distraktor/pengecoh.

2. KAJIAN LITERATUR

Validitas

Gronlund (2009: 70) menyebutkan bahwa validitas adalah ketepatan interpretasi yang diperoleh dari hasil penilaian. Lebih rinci Azwar (2010: 5) menyebutkan bahwa validitas berasal dari kata *validity* yang memiliki arti sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya. Namun secara khusus Allen & Yen (1979: 97) menyatakan bahwa validitas dari suatu perangkat tes dapat diartikan merupakan kemampuan suatu tes untuk mengukur apa yang seharusnya diukur.

Ada tiga tipe validitas, yaitu validitas isi, validitas konstruk dan validitas kriteria (Allen & Yen, 1979: 97). Selain itu, menurut (Azwar, 2011: 45-47) terdapat dua macam validitas isi, yaitu validitas kenampakan dan validitas logika. Validitas isi berarti sejauh mana suatu perangkat tes mencerminkan keseluruhan kemampuan yang hendak diukur (Azwar, 2011: 45), yang berupa analisis rasional terhadap domain yang hendak diukur. Validitas kenampakan didasarkan pada pertanyaan apakah suatu butir-butir dalam perangkat tes mengukur aspek yang relevan dengan domainnya. Validitas logika berkaitan dengan keseksamaan batasan pada domain yang hendak diukur, dan merupakan jawaban apakah keseluruhan butir merupakan sampel

representatif dari keseluruhan butir yang mungkin dibuat.

Validitas kriteria, disebut juga validitas prediktif, merupakan kesahihan suatu perangkat tes dalam membuat prediksi, dapat meramalkan keberhasilan siswa pada masa yang akan datang. Validitas prediktif suatu perangkat tes dapat diketahui dari korelasi antara perangkat tes dengan kriteria tertentu yang dikehendaki, yang disebut dengan variabel kriteria (Allen & Yen, 1979: 97; Azwar, 2011: 51).

Reliabilitas

Reynold (2006: 91) menyatakan bahwa reliabilitas mengacu pada kekonsistenan atau kestabilan hasil penilaian. Namun secara singkat Cohen (2007: 146) menyatakan bahwa reliabilitas sebagai kestabilan. Mengenai reliabilitas, Ebel & Frisbie (1991: 76), menyatakan bahwa jika tesnya memiliki konsistensi yang tinggi, maka tes tersebut akurat, *reproducible*, dan *generalizable* terhadap kesempatan testing dan instrumen yang sama.

Mehrens & Lehmann (1973: 249) menyatakan bahwa reliabilitas merupakan derajat keajegan (konsisten) di antara dua buah hasil pengukuran pada objek yang sama. Definisi ini dapat diilustrasikan dengan seseorang yang diukur tinggi badannya akan diperoleh hasil yang tidak berubah walaupun menggunakan alat pengukur yang berbeda dan skala yang berbeda. Kaitannya dengan dunia pendidikan, prestasi atau kemampuan seorang siswa dikatakan reliabel jika sudah dilakukan pengukuran. Kereliabelan ini bermakna hasil pengukuran akan sama informasinya, walaupun penguji berbeda, korektornya berbeda atau butir soal yang berbeda tetapi memiliki karakteristik yang sama.

Allen & Yen (1979: 62) menyatakan bahwa tes dikatakan reliabel jika skor

amatan mempunyai korelasi yang tinggi dengan skor yang sebenarnya. Selanjutnya dinyatakan bahwa reliabilitas merupakan koefisien korelasi antara dua skor amatan yang diperoleh dari hasil pengukuran menggunakan tes yang paralel. Dengan demikian, pengertian yang dapat diperoleh dari pernyataan tersebut adalah suatu tes itu reliabel jika hasil pengukuran mendekati keadaan peserta tes yang sebenarnya.

Oleh karena itu, dalam bidang pendidikan, pengukuran tidak dapat langsung dilakukan pada ciri atau karakter yang akan diukur. Ciri atau karakter ini bersifat abstrak. Hal ini menyebabkan sulitnya memperoleh alat ukur yang stabil untuk mengukur karakteristik seseorang (Mehrens & Lehmann, 1973: 103).

Berdasarkan uraian di atas, maka dalam pembuatan alat ukur dalam dunia pendidikan harus dilakukan secermat mungkin dan disesuaikan dengan kaidah-kaidah yang telah ditentukan oleh ahli-ahli pengukuran di bidang pendidikan. Kereliabilisan suatu alat ukur, yang berupa suatu indeks reliabilitas, dapat dilakukan penelaahan secara statistik. Nilai ini biasa dinamakan dengan koefisien reliabilitas (*reliability coefficient*). Menentukan nilai reliabilitas suatu tes dapat menggunakan aplikasi SPSS atau menghitung secara manual menggunakan formula sebagai berikut:

$$\hat{\alpha} = \frac{R}{R-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

dengan:

R : Banyaknya butir soal

σ^2 : Varians

Penetapan besarnya reliabilitas suatu instrumen tes dapat digunakan teknik korelasi dengan hasil korelasinya -1 sampai dengan +1. Tanda minus dan plus menunjukkan arah suatu hubungan tertentu. Angka

korelasi 1 atau korelasi sempurna, dalam bidang pendidikan, hampir tidak pernah tercapai. Umumnya selalu lebih kecil dari 1. Jika angka korelasinya diatas 0,60 dan kurang dari 1, maka instrumen tersebut memiliki korelasi tinggi atau reliabel, sedangkan jika angka korelasinya di bawah 0,50 ke bawah, maka instrumen tersebut berkorelasi rendah atau tidak reliabel (Basuki dan Haryanto, 2014: 144).

Selain validitas dan reliabilitas, dalam penelaahan soal secara kuantitatif terdapat tiga penelaahan butir soal yang didasarkan pada data empirik dari butir soal yang bersangkutan, yaitu Tingkat kesukaran, daya pembeda, dan analisis distraktor/pengecoh. Data empirik ini diperoleh dari soal yang telah diujikan.

Tingkat Kesukaran

Tingkat kesukaran butir soal adalah persentase atau proporsi dari peserta tes untuk menjawab benar suatu butir soal. Besarnya tingkat kesukaran berkisar 0,00 - 1,00. Semakin besar tingkat kesukaran yang diperoleh dari hasil hitungan, berarti semakin mudah soal itu dan soal itu harus direvisi. Suatu soal memiliki $p = 0,00$ artinya bahwa tidak ada siswa yang menjawab benar dan bila memiliki $p = 1,00$ artinya bahwa semua siswa menjawab benar.

keberagaman peserta tes terjadi ketika level optimum kesukaran adalah 0,50, yang mengindikasikan bahwa sebanyak 50% peserta tes menjawab benar, dan sebanyak 50% peserta tes menjawab salah. Akan tetapi, tidak setiap butir soal harus memiliki tingkat kesukaran 0,50. Boleh satu soal memiliki nilai p kurang dari 0,50, ataupun lebih besar dari 0,50, akan tetapi rata-rata dari keseluruhan p dari butir soal tetap 0,50.

Tingkat kesukaran butir soal memiliki 2 kegunaan, yaitu kegunaan bagi guru dan kegunaan bagi pengujian

dan pengajaran (Nitko, 2011: 303-305). Kegunaannya bagi guru adalah:

- a) Sebagai pengenalan konsep terhadap pembelajaran ulang dan memberi masukan kepada siswa tentang hasil belajar mereka,
- b) Memperoleh informasi tentang penekanan kurikulum atau mengecek terhadap butir soal yang bias.

Adapun kegunaannya bagi pengujian dan pengajaran adalah:

- a) Pengenalan konsep yang diperlukan untuk diajarkan ulang,
- b) Mengecek terhadap kelebihan dan kelemahan pada kurikulum sekolah,
- c) Memberi masukan kepada siswa,
- d) Mengecek kemungkinan adanya butir soal yang bias.

Daya pembeda

Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan siswa yang telah menguasai materi yang ditanyakan dan siswa yang tidak/kurang/belum menguasai materi yang ditanyakan. Manfaat daya pembeda butir soal adalah seperti berikut ini:

- a) Untuk meningkatkan mutu setiap butir soal melalui data empiriknya. Berdasarkan indeks daya pembeda, setiap butir soal dapat diketahui apakah butir soal itu baik, direvisi, atau ditolak.
- b) Untuk mengetahui seberapa jauh setiap butir soal dapat mendeteksi/membedakan kemampuan siswa, yaitu siswa yang telah memahami atau belum memahami materi yang diajarkan guru. Apabila suatu butir soal tidak dapat membedakan kedua kemampuan siswa itu, maka butir soal itu dimungkinkan seperti berikut ini:
 - (1) Kunci jawaban butir soal itu tidak tepat.

- (2) Butir soal itu memiliki 2 atau lebih kunci jawaban yang benar
- (3) Kompetensi yang diukur tidak jelas
- (4) Pengecoh tidak berfungsi
- (5) Materi yang ditanyakan terlalu sulit, sehingga banyak siswa yang menebak
- (6) Sebagian besar siswa yang memahami materi yang ditanyakan berpikir ada yang salah informasi dalam butir soalnya

Indeks daya pembeda setiap butir soal biasanya juga dinyatakan dalam bentuk proporsi. Semakin tinggi indeks daya pembeda soal berarti semakin mampu soal yang bersangkutan membedakan siswa yang telah memahami materi dengan siswa yang belum memahami materi. Indeks daya pembeda berkisar antara - 1,00 sampai dengan +1,00. Semakin tinggi daya pembeda suatu soal, maka semakin kuat/baik soal itu. Jika daya pembeda negatif (<0) berarti lebih banyak kelompok bawah (peserta didik yang tidak memahami materi) menjawab benar soal dibanding dengan kelompok atas (peserta didik yang memahami materi yang diajarkan guru).

Proses perhitungan daya pembeda dilakukan jika sudah dibedakan antara proses perhitungan daya pembeda kelompok kecil dengan kelompok besar. Terlebih dahulu kita menetapkan sampel yang mewakili kelompok pandai dan sampel yang mewakili kelompok kurang pandai, tetapi sebelumnya data telah diurutkan dari peserta didik yang memperoleh hasil yang tinggi ke peserta didik yang memperoleh hasil yang buruk. Kemudian sampel kelompok kecil dapat diambil 50% dari populasi untuk kelompok pandai dan kelompok kurang pandai, sedangkan untuk

kelompok besar masing-masing kelompok pandai dan kelompok kurang pandai diambil 27% dari populasi sehingga kelompok sedang sebanyak 54% dari populasi.

Analisis distraktor/pengecoh

Tes bentuk pilihan ganda terstruktur atas item permasalahan yang ditanyakan (atau pokok soal) dan option (atau sejumlah kemungkinan jawaban). Option sendiri dibagi menjadi dua, yaitu kunci jawaban dan pengecoh. Ini berarti, dari sekian jawaban yang disediakan hanya terdapat satu jawaban yang tepat yang disebut kunci jawaban, sedangkan kemungkinan jawaban yang lain (jawaban yang tidak benar) disebut pengecoh/distraktor.

Seseorang yang membuat soal pilihan ganda, terkadang ia tidak mudah untuk membuat pengecohnya. Kecilnya angka daya pembeda seringkali muncul karena pengecoh yang kurang berfungsi (Surapranata, 2009: 47).

Salah satu tujuan analisis distraktor adalah untuk memeriksa berapa banyak siswa pada kelompok atas dan kelompok bawah yang memilih tiap-tiap jawaban pada tes pilihan ganda. Oleh karena itu, dalam hal ini memuat dua pertanyaan. Pertama, apakah distraktor mengalihkan perhatian dari beberapa peserta tes? Jika tidak ada peserta tes memilih distraktor, ini berarti distraktor tidak berfungsi. Jika distraktor begitu jelas menyatakan sebagai jawaban tidak benar pada suatu tes namun tidak ada peserta tes yang memilih distraktor tersebut, berarti bahwa distraktor tersebut perlu direvisi atau diganti.

Pertanyaan kedua, apakah distraktor yang dibuat menarik lebih banyak dipilih oleh peserta tes kelompok bawah dari pada pada kelompok atas? Jika ya, berarti distraktornya baik. Distraktor atau

pengecoh berfungsi sebagai pengidentifikasi peserta tes yang berkemampuan tinggi. Pengecoh dikatakan berfungsi efektif jika banyak dipilih oleh peserta tes dari kelompok bawah, sebaliknya jika banyak dipilih oleh kelompok atas, maka distraktor/pengecoh tersebut tidak berfungsi sebagaimana mestinya.

3. HASIL DAN PEMBAHASAN

Validitas

• Validitas Isi

Validitas isi ditentukan menggunakan kesepakatan ahli (*expert*). Kesepakatan ahli bidang studi atau sering disebut dengan domain yang diukur menentukan tingkatan validitas isi. Hal ini dikarenakan instrumen pengukuran dibuktikan valid jika ahli meyakini bahwa bahwa instrumen tersebut mengukur penguasaan kemampuan yang didefinisikan dalam domain ataupun juga konstruk psikologi yang diukur (Retnawati, 2014: 7). Peneliti meminta kepada ahli untuk memeriksa ketepatan antara kesesuaian butir soal dengan indikator-indikatornya, redaksi penulisan soal, dan kesesuaian pilihan jawaban (pengecoh) pada pilihan ganda. Apabila masih ada kekeliruan dalam pembuatan instrumen, maka instrumen tersebut direvisi kembali.

Setelah dilakukan pemeriksaan oleh ahli, dalam hal ini sebagai validator, selanjutnya ahli memberikan penilaian terhadap instrumen. Penilaian tersebut terdiri dari 5 kriteria sebagai berikut:

Tabel 1. Kriteria Penilaian Butir Instrumen oleh Validator

Nilai	Keterangan
1	Tidak Relevan
2	Kurang Relevan
3	Cukup
4	Relevan
5	Sangat Relevan

Setelah diberikan penilaian oleh ahli, selanjutnya peneliti menghitung hasil penilaian menggunakan indeks validitas, diantaranya dengan indeks yang diusulkan oleh Aiken sebagai berikut.

$$V = \frac{\sum s}{N(c-1)}, \text{ dimana } s = r - l$$

Keterangan:

- r : rating penilai
 l : rating penilai kategori terendah
 c : kategori tertinggi
 N : jumlah penilai/responden
 (Aiken, 1980: 956)

Rentang angka V yang mungkin diperoleh adalah antara 0 sampai dengan 1. Semakin tinggi angka V (mendekati 1 atau sama dengan 1) maka nilai kevalidan sebuah item/butir soal juga semakin tinggi, dan semakin rendah angka V (mendekati 0 atau sama dengan 0) maka nilai kevalidan sebuah item/butir soal juga semakin rendah (Aiken, 1980: 957).

• Validitas Konstruk

Analisis untuk membuktikan validitas konstruk ialah menggunakan analisis faktor eksploratori. Analisis faktor eksploratori dapat dilihat dari persentasi varians yang dilihat dari nilai KMO (*Kaiser Meyer Olkin*). Nilai KMO dapat diperoleh melalui aplikasi SPSS IBM 20. Jika nilai KMO lebih dari 0,5, maka variabel dan sampel yang digunakan memungkinkan untuk dilakukan analisis lebih lanjut (Santoso, 2006: 22). Adapun langkah-langkah untuk mencari nilai KMO dengan SPSS IBM 20 adalah sebagai berikut:

- 1) Input data,
- 2) Klik *Analyze*,
- 3) Pilih *Dimension Reduction*, lalu pilih *Factor*,
- 4) Masukkan semua butir soal ke kolom *items*, lalu pilih *Descriptives* dan beri tanda centang pada *KMO and Bartlett's test of sphericity*, dan

- 5) Klik OK.

Estimasi Reliabilitas Instrumen

Teknik analisis data untuk estimasi reliabilitas instrumen ini menggunakan teknik estimasi konsistensi internal dengan formula *Chronbach-alpha* yang dibantu aplikasi SPSS IBM 20. Jika nilai *Cronbach's Alpha* 0,60 dan kurang dari 1, maka instrumen tersebut memiliki korelasi tinggi atau reliabel, sedangkan jika nilai *Cronbach's Alpha* di bawah 0,50 ke bawah, maka instrumen tersebut berkorelasi rendah atau tidak reliabel (Basuki dan Hariyanto, 2014: 105). Hal ini juga didukung oleh Surapranata (2009: 114), bahwa koefisien reliabilitas sebesar 0,5 dapat digunakan untuk tujuan penelitian.

Adapun langkah-langkah untuk mencari nilai *Cronbach's Alpha* dengan SPSS IBM 20 adalah sebagai berikut:

- 1) Input data,
- 2) Klik *Analyze*,
- 3) Pilih *Scale*, lalu pilih *Reliability Analysis*,
- 4) Masukkan semua butir soal ke kolom *items*, dan
- 5) Klik OK.

Tingkat Kesukaran Butir Soal

Teknik analisis data untuk tingkat kesukaran butir soal berupa pilihan ganda dapat dihitung dengan rumus:

$$p = \frac{\text{jumlah siswa yang menjawab benar}}{\text{jumlah siswa yang mengikuti testes}}$$

(Nitko, 2011: 301)

Tingkat kesukaran dalam bentuk soal uraian dapat dihitung dengan rumus:

$$p = \frac{\text{rata2 skor} - \text{skor minimum}}{\text{skor maksimum} - \text{skor minimum}}$$

(Nitko, 2011: 303)

Setelah dilakukan perhitungan, maka butir soal dapat dikategorikan menjadi butir soal yang sukar, sedang, dan mudah. Hal itu bergantung koefisien tingkat kesukarannya. Perhatikan Tabel berikut.

Tabel 2. Kategori Tingkat Kesukaran

Koefisien	Kategori
$p < 0,3$	Sukar
$0,3 \leq p \leq 0,7$	Sedang
$p > 0,7$	Mudah

(Surapranata, 2009: 21)

Daya Pembeda Butir Soal

Teknik analisis data untuk daya pembeda berupa pilihan ganda dapat dihitung dengan rumus:

$$DP = \frac{JB_A - JB_B}{n}$$

Dengan:

DP = Indeks Pembeda soal

JB_A = Jumlah peserta didik kelompok atas yang menjawab soal itu benar

JB_B = Jumlah peserta didik kelompok bawah yang menjawab soal itu benar

n = Persentase perbandingan ukuran kelompok.

(Nitko, 2011: 304)

Daya pembeda untuk soal uraian dapat dihitung dengan rumus;

$$DP = \frac{\bar{x}_{klmpk\ atas} - \bar{x}_{klmpk\ bawah}}{skor\ maksimum\ soal}$$

(Nitko, 2011: 304)

Setelah dilakukan perhitungan, maka butir soal dikategorikan menjadi butir soal yang diterima, direvisi, dan ditolak. Hal itu bergantung koefisien daya pembedanya. Jika ada soal yang ditolak, maka dapat dibuang atau diganti dengan butir soal yang baru. Perhatikan Tabel berikut:

Tabel 3. Kriteria Pemilihan Soal Berdasarkan Daya Pembeda

Kriteria	Koefisien	Keputusan
Daya Pembeda	$> 0,30$	Diterima
	$0,10 \text{ s.d } 0,29$	Direvisi
	$< 0,10$	Ditolak

(Surapranata, 2009: 47)

Kefektifan Distraktor/Pengecoh

Suatu distraktor yang efektif harus dipilih oleh beberapa peserta tes atau

minimal dipilih oleh 5% peserta tes (Basuki dan Hariyanto, 2014: 144). Pengecoh dikatakan berfungsi efektif jika banyak dipilih oleh peserta tes dari kelompok bawah, sebaliknya jika banyak dipilih oleh kelompok atas, maka distraktor/pengecoh tersebut tidak berfungsi sebagaimana mestinya.

4. KESIMPULAN

Validitas, reliabilitas, tingkat kesukaran, daya pembeda dan pengecoh (distraktor) merupakan hal yang penting yang harus diperhatikan oleh peneliti agar instrumen yang dibuat menjadi baik, sehingga bisa digunakan dalam dan menghasilkan data yang baik pula. 1) Validitas terbagi menjadi validitas isi dan konstruk. Validitas isi ditentukan oleh nilai V , sedangkan validitas konstruk ditentukan oleh nilai KMO. 2) Reliabilitas instrumen ditentukan oleh nilai *cronbach's alfa*. 3) Tingkat kesukaran butir soal ditentukan oleh nilai koefisien p . 4) Daya pembeda butir soal ditentukan oleh koefisien DB. 5) Distraktor/pengecoh ditentukan oleh persentase pemilihan dari peserta tes. Jika nilai atau persentase dari kelima hal ini berada pada kategori baik, maka instrumen yang dibuat peneliti memiliki kriteria yang baik.

5. REFERENSI

- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and psychological measurement*, 40(4), 955-959.
- Allen, M.j., & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey, Mexico: Brooks/Cole Publishing Company.
- Azwar, S. (2011). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Belajar.

- Basuki, I., & Hariyanto. (2014). *Asemen pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York: Routledge.
- Depdiknas. (2008). *Panduan penulisan butir soal*. Jakarta: Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah. Direktorat Pembinaan Sekolah Menengah Atas.
- Ebel, R.L. & Frisbie, D.A. (1991). *Essentials of educational measurement. Fifth edition*. New Delhi: Prentice Hall of India.
- Gronlund, N.E., Linn, R.L., & Miller, M.D. (2009). *Measurement & evaluation in teaching. Tenth edition*. New York: Macmillan Publishing Co., Inc.
- Mehrens, W.A. & Lehmann, I.J. (1973). *Measurement and evaluation in education and psychology*. New York : Hold, Rinehart and Wiston, Inc.
- Nitko, A.J. & Brookhart, S.M. (2011). *Educational assessment of students. Xth edition*. Upper Saddle River, New Jersey: Prentice Hall Englewood Cliffs.
- Retnawati, H. (2014). *Membuktikan validitas instrumen dalam pengukuran*. Diambil pada tanggal 8 Juli 2015 dari <http://www.evaluation-edu.com>.
- Reynolds, C.R, Livingston, R.B., & Willson, V. (2010). *Measurement and assessment in education. Second edition*. Upper Saddle River, New Jersey: Pearson Education.
- Santoso, S. (2006). *Seri Solusi Bisnis Berbasis TI: Menggunakan SPSS untuk Statistik Multivariat*. Jakarta: Elex Media Komputindo.
- Setyosari, P. (2013). *Metode penelitian pendidikan & pengembangan*. Jakarta: Prenadamedia group.
- Surapranata, S. (2009). *Analisis, validitas, reliabilitas, dan interpretasi hasil tes implementasi kurikulum 2004*. Bandung: PT. Remaja Rosdakarya.