

ISBN 978-602-72619-1-4

PROSIDING

SEMINAR NASIONAL PENDIDIKAN IPA VIII

“INTEGRASI SAINS-MORAL-SPIRITUAL (SMS)
DALAM PEMBELAJARAN IPA UNTUK MEWUJUDKAN
GENERASI EMAS INDONESIA
YANG BERTAQWA, MANDIRI, DAN CENDEKIA”



SABTU, 12 NOVEMBER 2016
RUANG SEMINAR FMIPA UNY



JURUSAN PENDIDIKAN IPA
FMIPA UNY

DAFTAR ISI

	Halaman
Halaman Judul	i
Kata Pengantar	ii
Sambutan Ketua Panitia	iii
Sambutan Dekan FMIPA UNY	v
Susunan Panitia Seminar	vii
Daftar Isi	ix
MAKALAH UTAMA	
AgusPurwanto, D.Sc	1
Alam sebagai Sumber Belajar IPAYang Mengintegrasikan Nilai Spiritual	
Prof. Dr. Zuhdan Kun Prasetyo, M.Ed.	33
Implementasi Pendidikan Karakter dalam Pembelajaran IPA	
MAKALAH PENDAMPING	
Rif'ati Dina Handayani, S.Pd, M.Si	44
Integrasi Budaya dalam Pembelajaran Sains	
Anatri Desstya, ST., M.Pd	50
Profil Keterampilan Proses Sains Guru-Guru di SD Negeri Pajang I Surakarta	
Apolonia Gerinus Gola, S.Pd	63
Integrated Science Nested Model untuk Meningkatkan Keterampilan Berpikir Kritis IPA Siswa SMP	
Suci Nurmatin, M.Pd	77
Potret Kemampuan Merancang Pembelajaran Calon Guru IPA Melalui Program Pengalaman Lapangan (PPL)	
Asmawati Amarullah	91
Pengembangan Perangkat Pembelajaran Pendekatan Lingkungan untuk Meningkatkan Literasi IPA Peserta Didik SMP	
AtinKurniawati, M.Pd	102
Paradigma <i>Project Based Learning</i> Menumbuhkan Kreativitas pada Mata Pelajaran IPA/Sains	
Dewiantika Azizah, S.Si., M.Pd	119
Implementasi Pendekatan Sainstifik Terintegrasi Nilai – Nilai Islami pada Pembelajaran Larutan Elektrolit dan Non Elektrolit	
Didik Setyawarno S.Pd. Si., M.Pd.	134
Analisis Butir Soal untuk Soal Pilihan Ganda Berdasarkan Teori Tes Klasik dengan Menggunakan Aplikasi SPSS.	
Elya Sumartik	145
Keefektifan Penerapan <i>Subject Specific Pedagogy (S Sp)</i> IPA Model <i>Guided Inquiry</i> untuk Meningkatkan Sikap Rasa Ingin Tahu dan Keterampilan Proses IPA Siswa SMP Kelas VII	



UNIVERSITAS NEGERI YOGYAKARTA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
JURUSAN PENDIDIKAN IPA

Alamat : Gedung D.07 Lantai 3 FMIPA UNY, Kampus Karangmalang, Yogyakarta 55281 Telp. 0274-586168 psw 442

SERTIFIKAT

3285/UN34.13/TU/2016

Diberikan kepada

Didik Setyawarno, S.Pd.Si., M.Pd.

sebagai

Pemakalah

dengan judul

**"Analisis Butir Soal untuk Soal Pilihan Ganda Berdasarkan Teori Tes Klasik
dengan Menggunakan Aplikasi SPSS"**

Seminar Nasional Pendidikan IPA VIII Tahun 2016

dengan tema "Integrasi Sains-Moral-Spiritual (SMS) dalam Pembelajaran IPA
untuk Mewujudkan Generasi Emas Indonesia yang Bertaqwa, Mandiri, Cendekia"
yang diselenggarakan di FMIPA Universitas Negeri Yogyakarta pada tanggal 12 November 2016

Yogyakarta, 12 November 2016
Ketua Jurusan Pendidikan IPA

Dekan FMIPA UNY

Dr. Hartono, M.Si.

NIP 19620329 198702 1 002

Dr. Dadan Rosana, M.Si.

NIP 19690202 199303 1 002

Analisis Butir Soal untuk Soal Pilihan Ganda Berdasarkan Teori Tes Klasik dengan Menggunakan Aplikasi SPSS

Didik Setyawarno

didiksetyawarno@uny.ac.id

Jurusan Pendidikan IPA, FMIPA UNY

Abstrak: Artikel ini bertujuan mengkaji tentang analisis butir soal dengan pendekatan teori tes klasik yang diperuntuk untuk guru atau mahasiswa calon guru baik bidang IPA atau Non-IPA. Aspek yang dikaji meliputi konsep dasar teori tes klasik yang dijabarkan menjadi berbagai formula penting dalam analisis butir soal yang meliputi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Selain itu, artikel juga mengkaji aplikasi SPSS untuk analisis butir soal bentuk pilihan ganda. Setelah membaca artikel ini, guru atau mahasiswa calon guru mempunyai pemahaman tentang konsep dasar teori tes klasik, validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Selain itu, guru atau mahasiswa calon guru mempunyai kemampuan menggunakan aplikasi SPSS untuk analisis butir soal bentuk pilihan ganda.

Kata Kunci: Teori Tes Klasik, Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, dan aplikasi SPSS.

PENDAHULUAN

Setiap proses pembelajaran yang telah dilakukan oleh guru selama waktu tertentu harus diketahui sejauhmana pembelajaran tersebut mampu meningkatkan kemampuan atau memberikan nilai tambah bagi siswanya. Tes merupakan salah satu alat untuk melakukan pengukuran, yaitu alat untuk mengumpulkan informasi karakteristik suatu objek. Objek ini bisa berupa kemampuan peserta didik, sikap, minat, maupun motivasi (Rosana, 2014:33). Tes tersebut dikalangan dunia pendidikan di sebut tes hasil belajar (THB).

Tes hasil belajar adalah salah satu alat ukur yang paling banyak digunakan untuk mengetahui hasil belajar seseorang dalam proses belajar-mengajar atau suatu program pendidikan. Istilah tes sering diartikan sebagai alat penilaian yang sifatnya

spesifik, dan secara sederhana diungkapkan dalam bentuk pertanyaan yang mengungkap tingkah laku, potensi maupun dikaitkan dengan hasil belajar (Suparwoto, 2005:4). Tes merupakan alat ukur yang standar dan obyektif sehingga dapat digunakan secara meluas untuk mengukur dan membandingkan keadaan psikis atau tingkah laku individu. Sebagai contoh, setiap tes yang diberikan oleh guru kepada siswa menuntut keharusan adanya respon dari subyek (orang yang dites) yang sedang dicari informasinya.

Sebagai alat untuk mengukur kemampuan siswa setelah mengikuti kegiatan pendidikan selama selang waktu tertentu, maka eksistensi tes menjadi sangat penting (Lababa, 2008:30). Sebuah tes yang baik, akan bisa mengungkapkan keadaan sebenarnya dari siswa, dan tes yang tidak baik tidak akan bisa mengungkap apa kemampuan sebenarnya siswa. Dengan demikian identifikasi terhadap setiap butir item soal perlu dilakukan sebelum butir soal tersebut digunakan dalam pengukuran kemampuan siswa. Identifikasi dan analisis butir item soal perlu dilakukan secara rutin oleh guru untuk perbaikan, pembenahan, dan penyempurnaan kembali terhadap butir-butir soal. Dengan cara ini, guru pada masa-masa yang akan datang tes hasil belajar yang disusun atau dirancang betul-betul dapat menjalankan fungsinya sebagai alat pengukur hasil belajar yang memiliki kualitas yang tinggi. Selain itu, analisis item butir soal dilakukan untuk mengetahui berfungsi atau tidaknya sebuah soal.

Analisis item butir soal pada umumnya dilakukan melalui dua cara, yaitu analisis kualitatif dan analisis kuantitatif (Subali, 2014). Analisis butir soal secara kualitatif dilaksanakan berdasarkan kaidah penulisan soal (tes tertulis, perbuatan, dan sikap). Penelaahan ini biasanya dilakukan sebelum soal digunakan atau diujikan. Aspek yang diperhatikan dalam penelaahan secara kualitatif mencakup aspek materi, konstruksi, bahasa atau budaya, dan kunci jawaban. Analisis item butir soal secara kuantitatif merupakan penelaahan butir soal didasarkan pada bukti empirik. Salah satu tujuan utama pengujian butir-butir soal secara empirik adalah untuk mengetahui sejauh mana masing-masing butir soal membedakan antara siswa yang tinggi

kemampuannya dengan siswa yang rendah kemampuannya.

Artikel ini akan memfokuskan pembahasan terkait dengan konsep dasar teori tes klasik yang dijabarkan menjadi berbagai formula penting dalam analisis butir soal yang meliputi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Selain itu, artikel juga mengkaji aplikasi SPSS untuk analisis butir soal bentuk pilihan ganda. Setelah membaca artikel ini, guru atau mahasiswa calon guru mempunyai pemahaman tentang konsep dasar teori tes klasik, validitas, reliabilitas, tingkat kesukaran, dan daya pembeda, serta mempunyai kemampuan menggunakan aplikasi SPSS untuk analisis butir soal bentuk pilihan ganda.

PEMBAHASAN

A. Teori Tes Klasik

Classical Test Theory yang dalam bahasa Indonesia sering disebut dengan teori tes klasik merupakan salah satu teori pengukuran yang tertua didunia pengukuran behavioral. Mamun Ali Naji Qasem menyatakan bahwa: *classical test theory introduces three concepts test score, true score, and error score. Within that theoretical framework, models of various forms have been formulated* (Qasem, 2013:78). Teori tes klasik memperkenalkan tiga konsep yaitu: skor tes, skor yang benar, dan skor galat. Model berbagai bentuk telah dirumuskan berdasarkan teori tersebut. Sebagai contoh, kita sering merujuk pada model uji klasik, yaitu model linear sederhana di mana postulat-postulas yang menghubungkan skor tes tampak (X) dengan jumlah dua variabel yang tidak dapat diamati, skor murni (T), dan skor kesalahan (E), yaitu, $X = T + E$. Ada tujuh macam asumsi yang ada dalam teori tes klasik ini. Allen & Yen (1979:57) menguraikan asumsi-asumsi teori klasik sebagai berikut.

1. Terdapat hubungan antara skor tampak (*observed score*) yang dilambangkan dengan huruf *X*, skor murni (*true score*) yang dilambangkan dengan *T* dan skor

kesalahan (error) yang dilambangkan dengan E . Dalam bahasa matematika dapat dilambangkan dengan $X = T + E$.

2. Skor murni (T) merupakan nilai harapan $e(X)$. Dengan demikian skor murni adalah nilai rata-rata skor perolehan teoretis sekiranya dilakukan pengukuran berulang-ulang (sampai tak terhingga) terhadap seseorang dengan menggunakan alat ukur.

3. Tidak terdapat korelasi antara skor murni dan skor pengukuran pada suatu tes yang dilaksanakan ($\rho_{et} = 0$).

4. Korelasi antara kesalahan pada pengukuran pertama dan kesalahan pada pengukuran kedua adalah nol ($\rho_{e1e2} = 0$).

5. Jika terdapat dua tes untuk mengukur atribut yang sama maka skor kesalahan pada tes pertama tidak berkorelasi dengan skor murni pada tes kedua ($\rho_{e1t2} = 0$).

Asumsi ini

6. Dua perangkat tes dapat dikatakan sebagai tes-tes yang paralel jika skor-skor populasi yang menempuh kedua tes tersebut mendapat skor murni yang sama ($T = T'$) dan varian skor-skor kesalahannya sama $\sigma_e^2 = \sigma_{e'}^2$.

7. Jika dua perangkat tes mempunyai skor-skor perolehan X_{t1} dan X_{t2} yang memenuhi asumsi 1 sampai 5 dan apabila untuk setiap populasi subyek $X_1 = X_2 + C_{12}$, dimana C_{12} adalah sebuah bilangan konstanta, maka kedua tes itu disebut tes yang paralel.

Asumsi-asumsi teori klasik sebagaimana disebutkan di atas memungkinkan untuk dikembangkan dalam rangka pengembangan berbagai formula yang berguna dalam melakukan pengukuran psikologis. Daya beda, indeks kesukaran, efektifitas distraktor, reliabilitas dan validitas adalah formula penting yang disarikan dari teori tes klasik (Lababa, 2008:30).

B. Validitas

Ellen A. Drost menyatakan bahwa *“validity is concerned with the*

meaningfulness of research components” (Drost, 2012:114). Dalam hal yang sama Ronald Jay Cohen menyatakan bahwa “*validity, as applied to a test, is a judgment or estimate of how well a test measures what it purports to measure in a particular context. More specifically, it is a judgment based on evidence about the appropriateness of inferences drawn from test scores* (Cohen, 2009:172). Pernyataan tersebut menyatakan bahwa validitas berkaitan dengan ketepatan keberartian komponen penelitian. Pengertian tersebut jika dikaitkan dengan butir soal sebagai alat ukur, maka validitas merupakan ketepatan alat ukur dengan hal yang diukur. Ada beberapa jenis validitas alat ukur, yaitu: validitas isi, validitas konstruksi, validitas ramalan, dan validitas sama saat. Validitas butir soal pilihan ganda/data dikotomi, jika butir soal di jawab benar bernilai 1 dan salah bernilai 0 di hitung dengan menggunakan teknik *Korelasi Point Biserial* (Koyan, 2012:56). Rumus yang digunakan untuk menghitung Korelasi Point Biserial sebagai berikut.

$$r_{pbi} = \frac{M_p - M_t}{s_t} \sqrt{\frac{p}{q}}$$

Keterangan:

r_{pbi} = koefisien korelasi *point biserial*

M_p = rerata skor dari subjek yang menjawab betul bagi butir yang dicari validitasnya

M_t = rerata skor total

s_t = standar deviasi dari skor total

p = proporsi peserta didik yang menjawab betul (banyaknya peserta didik yang menjawab betul dibagi dengan jumlah seluruh peserta didik)

q = proporsi peserta didik yang menjawab salah ($q = 1 - p$).

Suatu butir tes dinyatakan valid jika r hitung lebih besar daripada r tabel dengan taraf signifikansi atau taraf kekeliruan 5% ($r_{hit} > r_{tab}$ dengan taraf signifikansi 5%).

C. Reliabilitas

Reliabilitas berasal dari kata *reliability* yang dapat diartikan hal yang dapat dipercaya. Dalam hal yang sama, Drost menyatakan bahwa “*reliability is a major concern when a psychological test is used to measure some attribute or behaviour*” (Drost, 2012:106). Pengertian tersebut menyatakan bahwa reliabilitas adalah keterpercayaan, keterandalan, keajegan, konsistensi, atau kestabilan. Ada beberapa jenis reliabilitas, yaitu: (1) konsistensi internal, (2) stabilitas, dan (3) ekuivalen. Reliabilitas konsistensi internal alat ukur dapat dihitung dengan menggunakan rumus Koefisien Alpha-Cronbach, Kuder-Richardson (KR-20 atau KR-21), dan Teknik Belah Dua. Suparwoto menyatakan bahwa Koefisien AlphaCronbach dapat dimanfaatkan untuk analisis butir soal dengan skor benar +1 dan salah 0, atau dengan skor 1, 2, 3 berurutan dan cara ini merupakan upaya menetapkan koefisien reliabilitas instrumen/tes yang mengacu pada konsep *internal consistency* (Suparwoto, 2005:55). Rumus yang digunakan untuk menghitung Koefisien Alpha-Cronbach sebagai berikut.

$$r_{1.1} = \left(\frac{k}{k-1} \right) \left(\frac{SD_t^2 - \sum SD_i^2}{SD_t^2} \right)$$

Keterangan :

- $r_{1.1}$ = koefisien reliabilitas perangkat tes
 k = banyak butir tes
 SD_i^2 = varians skor tiap butir (item)
 SD_t^2 = varian skor total

Tingkat reliabilitas instrumen dapat diketahui dari nilai r sebagai berikut.

$r \leq 0,20 \Rightarrow$ sangat rendah

$0,20 < r \leq 0,40 \Rightarrow$ rendah

$0,40 < r \leq 0,60 \Rightarrow$ sedang

$0,60 < r \leq 0,80 \Rightarrow$ tinggi

$0,80 < r \leq 1,00 \Rightarrow$ sangat tinggi

D. Tingkat Kesukaran

Indeks kesukaran butir sebagaimana dinyatakan oleh Allen & Yen (1979:120) adalah “*proportion of examinees who get that item correct*”. Pernyataan tersebut menjelaskan bahwa tingkat kesukaran butir tes merupakan bilangan yang menunjukkan proporsi peserta ujian (*testee*) yang dapat menjawab betul butir soal tersebut. Sedangkan tingkat kesukaran perangkat tes adalah bilangan yang menunjukkan rata-rata proporsi *testee* yang dapat menjawab seluruh (perangkat) tes tersebut (Koyan, 2012:62). Rumus yang digunakan untuk menentukan tingkat kesukaran sebagai berikut.

$$P = \frac{nB}{n}$$

Keterangan:

- P = tingkat kesukaran butir tes
 nB = banyaknya subyek yang menjawab soal dengan betul
 n = jumlah subyek (*testee*) seluruhnya

Sebagaimana dinyatakan oleh Allen & Yen, tingkat kesukaran butir soal yang baik adalah 0,3 sampai 0,7. Butir dengan tingkat kesulitan dibawah 0,3 dianggap butir soal yang sukar sedangkan jika indeksnya diatas 0,7, butir soal tersebut dianggap mudah (Allen & Yen, 1979:121). Dengan demikian kriteria tingkat kesukaran (P) dapat dituliskan sebagai berikut.

0,00 - 0,30 => soal tergolong sukar

0,31 - 0,70 => soal tergolong sedang

0,71 - 1,00 => soal tergolong mudah

E. Daya Pembeda

Daya beda (diskriminasi) suatu butir tes adalah kemampuan suatu butir untuk

membedakan antara peserta tes yang berkemampuan tinggi dan berkemampuan rendah (Lababa, 2008:32). Pengertian tersebut menjelaskan bahwa daya beda butir tes merupakan kemampuan butir tes tersebut membedakan antara *testee* kelompok atas (tinggi) dan *testee* kelompok bawah (lemah). Rumus untuk menghitung daya beda butir tes adalah sebagai berikut.

$$D_B = \frac{nB_A}{n_A} - \frac{nB_B}{n_B}$$

Keterangan:

- nB_A = jumlah subyek yang menjawab betul pada kelompok atas
 nB_B = jumlah subyek yang menjawab betul pada kelompok bawah
 n_A = jumlah subyek kelompok atas
 n_B = jumlah subyek kelompok bawah

Kriteria Daya Beda (D_B) sebagai berikut.

- 0,40 – 1,00 => Soal baik
0,30 – 0,39 => Soal diterima dan diperbaiki
0,20 – 0,29 => Soal diperbaiki
0,00 – 0,19 => Soal ditolak

Jika “ D_B ” negatif, soal tersebut sangat buruk dan harus dibuang.

F. Analisis Butir Soal Pilihan Ganda dengan Aplikasi SPSS

Statistical Product and Service Solutions (SPSS) merupakan salah satu aplikasi statistic untuk mengolah data. Gaur menyatakan bahwa “*SPSS is a very powerful and user friendly program for statistical analyses*” (Gaur, 2006:15). Aplikasi tersebut sangat berguna dalam pengolahan data kuantitatif, tidak terkecuali dalam analisis butir soal baik untuk penentuan validitas, reliabilitas, tingkat Kesukaran, dan daya pembeda. Berikut cara analisis butir soal menggunakan aplikasi SPSS.

1. Pada *Variable View*, isikan variable semua butir soal yang akan dianalisis.

2. Tambahkan satu variabel jumlah di bawah variabel butir soal.
3. Pada *Data View*, isikan data butir soal yang akan di analisis (soal pilihan ganda: benar = 1 dan salah = 0).
4. Uji Validitas:
 - a. Pilih menu *Analyze --> Correlate --> Bivariate*
 - b. Masukkan variabel soal dan jumlah ke Kotak "*Variables*" pada kotak dialog *Bivariate Correlations*.
 - c. Pilih *Pearson, Two-tailed*, dan *Flag significant correlation* dengan mengklik tombol tersebut kemudian Klik OK.
 - d. Pembacaan Hasil: pada kolom "jumlah" dengan ketentuan jika nilai sig < 0,05 maka butir soal tersebut valid, dan sebaliknya jika nilai sig \geq 0,05 maka butir soal tersebut tidak valid.
5. Uji Reliabilitas:
 - a. Pilih menu *Analyze --> Scale --> Reliability Analysis*.
 - b. Masukkan semua variabel soal yang akan di analisis ke Kotak "*Items*"
 - c. Pada kotak dialog *Reliability Analysis* untuk kolom Model pilih *Alpha*, klik tombol *Statistics*, dan pilih Item selanjutnya *Continue* dan OK.
 - d. Klik OK.
 - e. Pembacaan Hasil: pada tabel *Reliability Statistics* akan muncul nilai *Cronbach's Alpha* yang menyatakan koefisien reliabilitas butir soal tersebut.
6. Uji Tingkat Kesukaran
 - a. Pilih menu *Analyze --> Descriptive Statistics --> Frequencies*.
 - b. Masukkan semua variabel butir soal ke kotak "Variable(s)"
 - c. Pada kotak dialog *Frequencies*, pilih tombol *Statistics*, klik *Mean*, dan Klik *Continue* dan OK.
 - d. Pembacaan Hasil: pada tabel *statistics* muncul nilai MEAN untuk setiap butir soal dengan ketentuan sebagai berikut.
0,00 - 0,30 => soal tergolong sukar
0,31 - 0,70 => soal tergolong sedang

0,71 - 1,00 => soal tergolong mudah

7. Uji Daya Pembeda

Penentuan daya pembeda setiap butir soal dengan melihat nilai *person correlations* pada hasil uji validitas yang telah dilakukan dengan ketentuan sebagai berikut.

0.40 – 1.00 => Soal baik

0.30 – 0.39 => Soal diterima dan diperbaiki

0.20 – 0.29 => Soal diperbaiki

0.00 – 0.19 => Soal ditolak

PENUTUP

Asumsi-asumsi teori klasik merupakan dasar pengembangan berbagai formula yang berguna dalam melakukan pengukuran butir soal mencakup daya beda, indeks kesukaran, efektifitas distraktor, reliabilitas dan validitas. Validitas butir soal merupakan ketepatan butir soal tersebut dengan aspek yang diukur dalam penilaian hasil belajar. Teknik Korelasi Point Biserial merupakan salah satu teknik untuk menentukan butir soal tersebut valid atau tidak. Reliabilitas adalah keterpercayaan, keterandalan, kejelasan, konsistensi, atau kestabilan. Reliabilitas butir soal dapat dihitung salah satunya dengan menentukan koefisien Alpha-Cronbach. Tingkat kesukaran butir tes merupakan bilangan yang menunjukkan proporsi peserta ujian (testee) yang dapat menjawab betul butir soal tersebut. Daya beda (diskriminasi) suatu butir tes adalah kemampuan suatu butir untuk membedakan antara peserta tes yang berkemampuan tinggi dan berkemampuan rendah. Aplikasi SPSS dapat digunakan untuk analisis butir soal dengan pendekatan teori klasik tes.

DAFTAR PUSTAKA

- Allen & Yen. 1979. *Introduction to Measurement Theory*. Belmont, California: Wadsworth, Inc.
- Arif Pratisto. 2004. *Cara Mudah Mengatasi Masalah Statistik dan Rancangan Percobaan dengan SPSS*. Jakarta: Gramedia.
- Bambang Subali. Analisis Soal Baik Kualitatif Maupun Kuantitatif. *Disajikan pada Kegiatan Workshop Item Development Dosen Poltekkes Kebidanan Politeknik Kesehatan Surakarta tanggal 18-19 Agustus 2014 di Griya Persada Conventional Hotel & Resort, Jl Boyong*

Kaliurang Barat.

- Carmine, EG & Zeller, RA. 1979. *Reliability and Validity Assessment*. Beverly Hills, California: Sage Publications, Inc.
- Cohen, R.J. 2009. *Psychological Testing and Assessment: An Introduction to Tests and Measurement 7th Edition*. New York, USA: Mc-Graw-Hill.
- Dadan Rosana. 2014. *Evaluasi Pembelajaran Sains*. Yogyakarta: UNY Press.
- Drost, E.A. Validity and Reliability in Social Science Research. *Education Research and Perspectives, Vol.38, No.1, 2012*.
- Gaur, A.S. 2006. *Statistical Methods for Practice and Research*. Delhi, India: SAGE Publications.
- I Wayan Koyan. 2012. *Konstruksi Tes*. Bali: Undiksha Press.
- Junaidi Lababa. Analisis butir soal dengan teori tes klasik: Sebuah pengantar. *Jurnal Iqra' Volume 5 Januari - Juni 2008*.
- Qasem, M.A.N. A Comparative Study of Classical Theory (Ct) and Item Response Theory (Irt) In Relation To Various Approaches of Evaluating the Validity and Reliability of Research Tools. *IOSR Journal of Research & Method in Education (IOSR-JRME) Volume 3, Issue 5, pp 77-81, 2013..*
- Suparwoto. 2005. *Penilaian Pencapaian Hasil Belajar Fisika*. Yogyakarta: FMIPA UNY.
- Zoghi, M and & Valipour, V. A Comparative Study of Classical Test Theory And Item Response Theory In Estimating Test Item Parameters In A Linguistics Test. *Indian Journal of Fundamental and Applied Life Sciences, Vol. 4 (S4), pp. 424-435, 2014*.