

Pendeteksian DIF pada Perangkat Tes Objektif Penilaian Akhir Semester IPA dengan Menggunakan Permodelan Rasch

M. Ardi Setiawan¹, Purwo Susongko², Muriani Nur Hayati³

^{1,2,3}Prodi Pendidikan IPA, FKIP Universitas Pancasakti Tegal, Indonesia

Korespondensi. E-mail: setiawnardi129@gmail.com

Abstrak

Tujuan penelitian ini adalah bagaimana karakteristik butir soal Penilaian Akhir Semester (PAS) dan mendeteksi ada tidaknya *Differential Item Functioning* (DIF) pada perangkat tes objektif Penilaian Akhir Semester (PAS) ganjil kelas VII IPA di SMP Negeri Kabupaten Tegal. Populasi yang digunakan seluruh SMP Negeri di Kabupaten Tegal. Metode yang digunakan ialah Pengumpulan data dengan metode dokumentasi. Teknik analisis penelitian ini menggunakan permodelan *Rasch* dengan uji Wald. Berdasarkan hasil analisis karakteristik butir pada perangkat tes dan pendeteksian DIF dengan menggunakan uji wald menunjukkan bahwa terdapat karakteristik berupa perbedaan tingkat kesukaran butir soal penilaian akhir semester mata pelajaran IPA yang diujikan pada peserta didik serta pada perangkat soal Penilaian Akhir Semester (PAS) Ganjil kelas VII SMP Negeri di Kabupaten Tegal tahun ajaran 2019/2020 terdapat 2 butir yang terdeteksi mengandung DIF berdasarkan gender yaitu butir nomor 2 dan 9.

Kata Kunci: *Differential Item Functioning* (DIF), Penilaian Akhir Semester, *Rasch*

DIF Detection on Objective Test Devices IPA Semester Final Assessment with the use of Rasch Modelling

Abstract

The purpose of this research is how the particulars characteristic of the final assessment (PAS) and detecting the absence of Differential Item Functioning (DIF) in the objective test device the final assessment Semester (PAS) Odd class VII IPA in SMP Negeri Tegal district. The population used in the whole state junior high School in Tegal. Method used is data collection with documentation method. The analysis technique uses Rasch modelling with Wald test. Based on the results of the characteristic analysis of grain on test devices and the DIF detection using the Wald test, indicating that there are differences in the difficulty level of grains problem of final assessment of the semester of SCIENCE subjects that are tested on the students and on the device in the end of Semester assessment (PAS) Odd class VII SMP Negeri in Tegal District School 2019/2020 There are 2 and 9 grains detected containing DIF based.

Keywords: *Differential Item Functioning, final assessment of semester, Rasch model*

PENDAHULUAN

Kabupaten Tegal saat ini dalam jenjang pendidikan menengah pertama sudah menggunakan kurikulum 2013 yang telah direvisi setiap tahunnya. Proses yang dilakukan untuk mencapai keberhasilan tujuan dari kurikulum K- 13 atau untuk mengetahui keberhasilan dari proses pembelajaran salah satunya dengan evaluasi, dengan evaluasi dapat diketahui sejauh mana tujuan pembelajaran tercapai. Pembelajaran merupakan bagian proses belajar mengajar dalam suatu interaksi sosial guna mencapai tujuan pembelajaran. Hal tersebut dapat melalui ulangan harian, tugas, penilaian tengah semester serta penilaian akhir semester.

Penilaian akhir semester merupakan salah satu kegiatan dalam jenjang satuan pendidikan menengah pertama dalam mengevaluasi hasil belajar peserta didik selama satu semester. Hal ini mengukur kemampuan peserta didik dalam kemampuan kognitif yang telah diajarkan guru selama proses pembelajaran terkait dengan materi – materi yang diajarkan.

Menurut Susongko (2013) penilaian adalah suatu usaha untuk mendapatkan berbagai informasi secara berkala, berkesinambungan, dan menyeluruh tentang proses dan hasil dari pertumbuhan dan perkembangan yang telah dicapai oleh peserta didik melalui program kegiatan belajar.

Susongko (2013) menyatakan bahwa bentuk tes terbagi atas dua bentuk yaitu tes uraian (constructed response test) dan tes obyektif (selected response test). Dalam pengembangan teori tes modern menggunakan *Rasch* models, menurut Sumintono & Widhiarso (2015) pemodelan *Rasch* memiliki keunggulan dibanding metode lainnya, khususnya teori tes klasik, yaitu kemampuan melakukan prediksi terhadap data hilang (*missing data*), berdasarkan pola respon individu. Perbedaan tersebut dapat menyebabkan respon peserta didik dalam menjawab soal penilaian akhir semester pun

berbeda. Perbedaan tersebut akan menyebabkan peserta didik dengan kemampuan yang sama, dan pada butir soal yang sama memperoleh peluang yang berbeda dalam menjawab benar atau biasa disebut dengan *Differential Item Functioning* (DIF).

Keberadaan DIF merupakan hal yang penting bagi suatu perangkat tes karena dapat menunjukkan apakah suatu tes adil bagi semua kelompok peserta didik atau tidak. DIF ada dua jenis DIF internal dan eksternal, sedangkan untuk melakukan pengukuran pada perangkat tes objektif dengan jenis DIF eksternal. Diferensial Item Functioning (DIF) merupakan ancaman potensial terhadap validitas tes. Standar APA (American Psychological Association) dan ITC (International Test Commission) menekankan perlunya memeriksa DIF untuk menjamin penggunaan tes yang adil. (AERA, APA, NCME, 1999; COP-ITC, 2000; Hernández 2003).

Differential Item Functioning (DIF) digunakan untuk mendeteksi adanya perbedaan karakteristik antara suatu kelompok dengan kelompok yang lain. Apabila sebuah perangkat soal terdeteksi DIF dapat menyebabkan perangkat tes tersebut tidak fair karena menguntungkan salah satu kelompok (Holland & Wainer, 1995). Sehingga dalam sebuah perangkat tes sebaiknya tidak terdeteksi mengandung DIF.

Secara konseptual, suatu butir menunjukkan DIF jika peserta tes memiliki kemampuan sama berada dalam kelompok berbeda, tidak mempunyai probabilitas sama untuk menjawab betul (Hambleton dan Swaminthan, 1985). Untuk mendeteksi apakah suatu butir mengandung DIF atau tidak, diperlukan indeks DIF, yaitu indeks yang menunjukkan sekuat indikasi DIF pada butir soal. Jika tingkat indikasi DIF secara praktik signifikan, dapat dengan mengujinya memakai uji statistik tertentu atau hanya dengan indeksinya saja, maka butir soal yang bersangkutan dikatakan terdeteksi DIF.

Permasalahan yang terungkap adalah hasil Penilaian Akhir Semester yang berbeda – beda di setiap sekolah, dapat dipengaruhi oleh perbedaan latar belakang peserta didik, faktor bias item (bias butir) pada soal Penilaian Akhir Semester (PAS) dan belum banyak yang melakukan penelitian tentang bias butir pada Penilaian Akhir Siswa (PAS).

Pengembangan dan mengevaluasi tes merupakan pertimbangan yang paling pokok atau validitas tes dalam penggunaan perangkat dari tes Penilaian Akhir Semester. Suwanto (2012) menyatakan bahwa Proses validasi dimulai dari pengajuan sebuah pernyataan yang eksplisit mengenai interpretasi – interpretasi dari suatu skor atau nilai tes. Sehingga terjadi pada skor atau nilai tes yang berbeda – beda akibat validits perangkat dalam pengukuran bias pada setiap butir tesnya.

Bias butir dalam suatu pengukuran mengindikasikan adanya kesalahan sistematika dalam pengukuran tersebut. Ada dua macam bias, yaitu bias eksternal dan bias internal. Bias eksternal merupakan bias yang menitikberatkan pada validitas kontrak dan validitas prediktif dari suatu tes secara keseluruhan. Bahwa bias internal yang juga biasa disebut dengan bias butir merupakan aspek dari bias dalam tes yang berkaitan dengan sifat-sifat psikometri dari suatu butir tes dan tes secara keseluruhan. (Retnawati, 2014).

Ada beberapa metode untuk mendeteksi DIF dalam suatu perangkat tes. Dari pendeteksian dengan pendekatan klasik hingga pendeteksian dengan pendekatan modern. Metode Wald test salah satu metode yang digunakan pendeteksi DIF yaitu Bagaimana karakteristik butir soal Penilaian Akhir Semester(PAS), butir-butir mana yang mengandung DIF pada soal Penilaian Akhir Semester, dan bagaimana *confidence plot* butir soal Penilaian Akhir Semester (PAS).

METODE

Penelitian ini menggunakan pendekatan kuantitatif, sedangkan jenis penelitian yang digunakan penelitian deskriptif. Penelitian

deskriptif merupakan penelitian yang berusaha mendeskripsikan suatu gejala, peristiwa, kejadian yang terjadi saat sekarang.(Susongko, 2014).

Penelitian deskriptif yang digunakan adalah analisis dokumenter karena penelitian ini menggambarkan serta menyimpulkan suatu obyek yang diperoleh dengan cara mengumpulkan data untuk mendapatkan pemecahan masalah dari masalah yang diteliti.

Analisis dokumenter menurut Susongko (2014) didefinisikan sebagai suatu teknik riset yang bersifat obyektif, sistematis, dan deskripsi kuantitatif makna tersembunyi dari komunikasi. Teknik Analisis data dalam pendeteksian DIF dengan menggunakan uji Wald yang akan menghasilkan *output p-value* dan nilai *z*. Pada *R Programming version 4.0.0* terdapat pada paket *eRm*. Selain uji Wald pada paket *eRm* juga dapat mendeteksi DIF dengan menggunakan *confidence plot*.

HASIL DAN PEMBAHASAN

Pendeteksian *Differential Item Functioning* (DIF) dengan Uji Wald, plot DIF, dan kurva karakteristik. Berikut hasil dari pendeteksian DIF:

Wald test

Pendeteksian DIF dilakukan dengan melihat nilai peluang dengan taraf signifikansinya. Taraf signifikansi yang digunakan dalam penelitian ini adalah 1% atau 0,01. Hipotesis yang digunakan dalam uji Wald sebagai berikut:

H0 : tidak ada pengaruh jenis kelamin secara simultan terhadap proporsi menjawab benar soal

H1 : ada pengaruh jenis kelamin secara simultan terhadap proporsi menjawab benar soal

H0 akan ditolak jika $p\text{-value} < 0,01$. Hasil analisis dengan uji wald dapat dilihat pada Tabel 1.

Tabel 1. Hasil analisis DIF

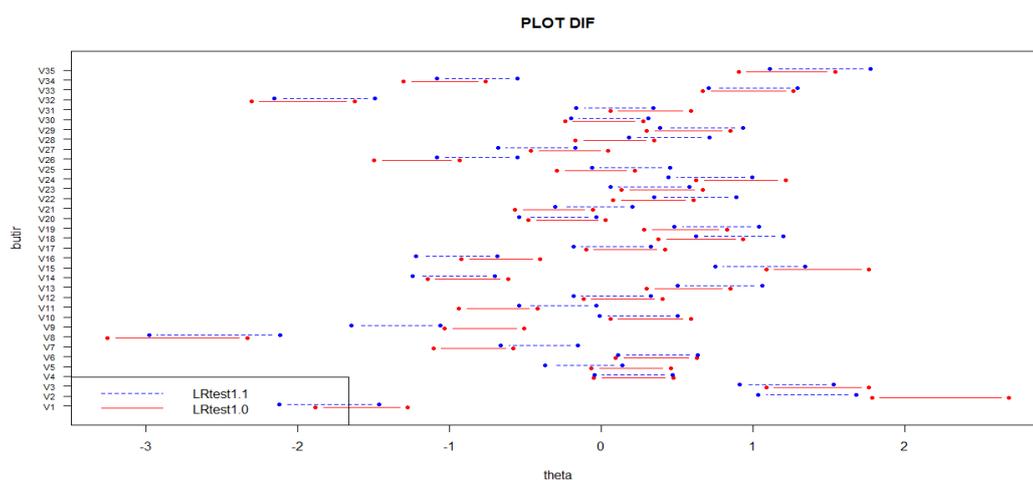
Butir	z-statistic	p-value	Keterangan
2	-3.094	0.002	DIF
9	-2.914	0.004	DIF

Berdasarkan Tabel 1, hasil pendeteksian DIF menggunakan uji Wald menunjukkan bahwa pada taraf signifikansi 0,01 terdapat 2 butir soal yang terdeteksi

mengandung DIF yakni butir nomor 2, dan 9 Adapun terdapat 33 butir soal yang tidak mengandung DIF.

Plot DIF

Pada hasil analisis menggunakan *R Programming version 4.0.0*, plot DIF yang dihasilkan berupa *Confidence plot*. *Confidence plot* tersebut menentukan butir-butir mana yang terdeteksi DIF.



Gambar 1. Plot DIF pada perangkat tes PAS IPA
Sumber: R Program 4.0.0

Garis LR test1.1 menunjukkan peserta didik laki-laki, sedangkan garis LRtest1.0 menunjukkan peserta didik perempuan. Akan terdeteksi DIF apabila garis LRtest 1.1 dengan garis LRtest1.0 tidak berhimpit sedangkan apabila garis LRtest1.1 dengan garis LRtest1.0 berhimpit maka tidak terdeteksi DIF.

Berdasarkan data yang dianalisis sesuai karakteristik dari masing – masing butir dalam penilaian akhir semester IPA 2019/2020. Sesuai kisi – kisi dari MGMP kab. Tegal terdapat 6 bab dan 40 indikator dengan 35 indikator untuk tes objektif / pilihan ganda. Jenis – jenis indikator yang digunakan sesuai dengan taksonomi Bloom yakni dari C2 – C6 dan sesuai anjuran dari pemerintah dalam proses pembelajaran dengan menggunakan pendekatan scientific literacy atau inkuiri (Kemedikbud, 2017), sehingga menyesuaikan masing – masing indikator pada taksonomi

Bloom pada perangkat kerja atau perangkat tes.

Menurut Wenning dan Vierya (2015) mengembangkan kerangka kerja yang cukup lengkap berkaitan hubungan kemampuan proses intelektual dan praktis ilmiah yang digolongkan ke dalam tingkat capaian level intelektual dan berhubungan dengan tingkat inkuiri. Selanjutnya tiap – tiap level terhubung ke taksonomi Bloom sesuai tujuan pendidikan untuk membuktikan kesesuaian kategori (Wenning, 2007). Penilaian Akhir Semester (PAS) IPA di kabupaten Tegal dengan menyesuaikan silabus dari pemerintah serta indikator – indikator yang tercantum dalam kisi – kisi PAS IPA 2019/2020 guna membuktikan kesesuaian kategori untuk tujuan pendidikan. Hayati (2017) menyatakan bahwa pendidikan "Sains untuk semua" perlu dipersiapkan, disetujui, dan diadaptasi ke dalam masyarakat, karena pendidikan sains

akan membantu mereka meningkatkan pemahaman dan kebiasaan berpikir mereka yang diperlukan untuk dapat berpikir kritis, tanggap dan peduli, mandiri, dan positif untuk pembangunan bangsa Indonesia.

Penilaian akhir semester salah satu alat evaluasi pendidikan disetiap semester untuk menguji kesesuaian kategori mealalui taksonomi Bloom. Terdapat 40 butir tes diantaranya 35 butir tes objektif dan 5 butir tes subjektif/essay, pada 35 butir tes objektif dimana laki – laki dan perempuan memiliki tingkat kesukaran dalam menegrjakan butir tes penilaian akhir semester IPA yakni 51 – 57 persen. Butir tes yang digunakan pada penilaian akhir semester ini tipe soal C2 – C6, terlebih untuk tipe soal C3 - C6 tergolong sukar bagi peserta didik dari data yang sudah dianalisis (tabel 4.1). Menurut Bloom, item yang bertanya tentang menghafal dikategorikan sebagai tingkat kemampuan kognitif terendah. Oleh karena itu item yang mengukur proses ini cenderung memiliki tingkat kesulitan yang rendah. Semakin tinggi tingkat proses kognitif yang dilakukan, semakin tinggi tingkat kesulitan soal pertanyaan yang mengukurnya (Suminto,2018). Tingkat proses kognitif yang dikembangkan oleh Bloom bergerak dari ingatan, pemahaman, penerapan, analisis, evaluasi, dan akhirnya sintesis.

Selanjutnya, dalam dari data yang dihimpun lalu dianalisis pendeteksian DIF dengan menggunakan model Rasch (tabel 4.2) menemukan adanya 2 butir soal yang terdeteksi mengandung DIF berdasarkan perbedaan gender yaitu butir 2, dan 9. Terdeteksinya DIF menunjukkan perbedaan gender mempengaruhi respon jawaban pada peserta didik yang memiliki kemampuan sama.

Keberfungsian butir pada perangkat tes/penilaian akhir semester IPA merupakan Analisis DIF diperlukan untuk memeriksa parameter atau membangun invarian (Zumbo, 2007). Invariansi berkaitan dengan respons item yang harus dikondisikan secara

independen pada profil atribut. Oleh karena itu, analisis DIF penting untuk menyelidiki invarians interaksi atribut-item lintas kelompok (Hou dkk, 2014). Prosedur ini mengidentifikasi kelompok orang yang melanggar invariansi pengukuran karena DIF dalam model Rasch dikotomis. (Komboz, B., Strobl, C., & Zeileis, A. (2018).

DIF terdeteksi pada butir soal nomor 2, yang dikategorikan butir soal sukar karena pada peserta didik laki-laki memiliki tingkat kesukaran sebesar 1.358 Terlalu Sukar untuk laki – laki, dengakan perempuan tingkat kesukaran sebesar 2.236 atau terlalu sukar. Butir soal ini menguntungkan peserta didik laki-laki pada interval tingkat kemampuan 0,9 sampai dengan 1,7 dan menguntungkan peserta didik perempuan pada interval tingkat kemampuan 1,9 sampai dengan 2,3. Artinya pada interval tingkat kemampuan tersebut peserta didik memberikan respon jawaban benar.

2. Perhatikan tabel di bawah ini!

No.	Besaran	Satuan	Alat ukur
1.	Suhu	Kelvin	Termometer
2.	kuat arus	amperemeter	ampere
3.	panjang	meter	mistar
4.	massa	kilogram	stopwatch

Besaran pokok dengan satuan SI serta alat ukurnya yang benar adalah

A. 1 dan 2
B. 1 dan 3
C. 2 dan 3
D. 2 dan 4

Gambar 2. Butir tes no 2 pada perangkat tes PAS IPA

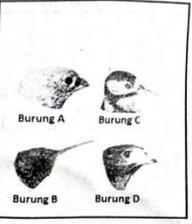
Sumber: Perangkat Tes PAS IPA 2019/2020

Dilihat dari gambar 12 bahwa kontruksi soal merupakan soal tentang pengukuran namun dengan tipe soal analisi yang menunjukkan tabel dimana siswa diminta untuk mencocokkan satuan internasional (SI) dengan alat ukur. Pada soal nomor 2 menunjukkan jawaban yang benar ialah C (2 dan 3), kesukaran pada tipe butir ini tinggi karena keliruan antara amperemeter dan amper sehingga banyak juga peserta didik menjawab A (1 dan 2) dan B (1 dan 3). Dalam kisi – kisi yang digunakan “ Disajikan tabel data, siswa dapat menganalisis Besaran pokok

dengan satuan SI serta alat ukurnya yang benar”, tipe soal IPA dengan bab fisika sangat memberi peluang kepada laki – laki dalam mengerjakan ketimbang perempuan dilihat dari tingkat kesukaran butir. Dalam penelitian Nur Triyono (2014) menunjukkan bahwa peluang mudah mengerjakan pada soal fisika diuntungkan kepada laki – laki dan dianggap sukar bagi perempuan.

9. Gunakan gambar paruh burung dan kunci determinasi berikut

Kunci Determinasi Sederhana	
1a Paruh pendek dan kuat	2
1b Paruh panjang dan runcing.....	3
2a Paruh tajam dan melengkung.....	Elang
2b Paruh tidak melengkung	Pipit
3a Berbentuk kantong.....	Pelikan
3b Tidak berbentuk kantong	4
4a Paruh kuat	Pelatak
4b Paruh lembut	Kolibri



Burung A teridentifikasi sebagai ...

- burung elang
- burung pipit
- burung pelatak
- burung kolibri

Gambar 3. Butir tes no 9 pada perangkat tes PAS IPA

Sumber: Perangkat Tes PAS IPA 2019/2020

Dilihat dari gambar 13 bahwa konstruksi soal merupakan soal tentang klasifikasi makhluk hidup dengan model soal disajikan data kunci determinasi dimana siswa diharuskan menganalisis soal tersebut. Pada soal nomor 9 menunjukkan jawaban yang benar ialah B (burung pipit), kesukaran pada tipe butir ini tinggi karena berbentuk analisis yang disajikan dalam bentuk kunci determinasi serta soal ini memang tergolong sulit. Dalam kisi – kisi yang digunakan “Disajikan gambar dan kunci determinasi , siswa dapat menentukan jenis hewan berdasarkan kunci determinasi tersebut”, tipe soal IPA dengan bab biologi terdapat perbedaan peluang dalam mengerjakan (gambar 11) antara laki – laki dan perempuan, sehingga pada butir tes nomor 9 terjadi DIF. Hal ini dikarenakan tingkat kesukaran butir antara laki – laki (terlalu mudah) dan perempuan (mudah)serta tingkat kepercayaan diri dalam siswa yang berbeda – beda sehingga mempengaruhi dalam mengerjakan soal.

SIMPULAN

Hasil analisis karakteristik butir dan pendeteksian DIF menggunakan uji wald menunjukkan bahwa terdapat karakteristik berupa perbedaan tingkat kesukaran butir soal penilaian akhir semester mata pelajaran IPA yang diujikan pada peserta didik serta pada perangkat soal penilaian akhir semester mata pelajaran IPA kelas VII SMP Negeri di kabupaten Tegal tahun ajaran 2019/2020 terdapat 2 butir soal yang terdeteksi mengandung DIF yaitu butir nomor 2 dan 9.

Bagi penyusun soal sebaiknya menggunakan soal-soal yang sesuai dengan kemampuan peserta didik dan yang secara empiris terbukti berkualitas baik dan tidak memuat DIF. Bagi penelitian lain perlu adanya penelitian lebih lanjut dengan menggunakan data dan metode yang berbeda untuk memperluas kajian tentang DIF.

DAFTAR PUSTAKA

- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999). Standards for educational and psychological testing. Washington DC: American Educational Research Association.
- Colegio Oficial de Psicólogos (COP) and Comisión Internacional de Tests(ITC) (2000). Directrices internacionales para el uso de los tests. Madrid: (International Guidelines for test use). Madrid: COP.
- Egiseptifera, Fifi. 2016. *Keberfungsian Butir Diferensial pada Ujian Akhir Semester Matematika di Sekolah Menengah Pertama (Studi Deskriptif Analisis Dokumenter pada Perangkat Ujian Akhir Semester Ganjil Kelas VII SMP di Lingkungan Kabupaten Tegal Tahun Ajaran 2015/2016)*. Skripsi. Pendidikan Matematika. Fakultas Keguruan dan Ilmu Pendidikan. Universitas Pancasakti Tegal.
- Hambleton, R. K. Swaminathan, H. (1985). *Item Respon Theory. Participle and Application*. Boston : Kluwer – Nijhoff Publishing.
- Hambleton, R. K., Swaminathan. H., & Rogers, H. J. (1991) *Fundamentals of*

- item response Theory. California: Sage Publications, *The International Professional Publishers*.
- Hayati, MN. (2017). The Use Of Science Literacy Taxonomy To Measure Chemistry Literacy Of The Science Teacher Candidates. *Unnes Science Education Journal*, 6(1).1496 – 1502
- Hou, L., la Torre, J. D., & Nandakumar, R. (2014). Differential item functioning assessment in cognitive diagnostic modeling: Application of the Wald test to investigate DIF in the DINA model. *Journal of Educational Measurement*, 51(1), 98-125.
- Lederman, N.G., & Lederman, J.S. (2004). Revising instruction to teach nature of science. *The Science Teacher*, 71(9). 36 – 39.
- Komboz, B., Strobl, C., & Zeileis, A. (2018). Tree-based global model tests for polytomous Rasch models. *Educational and Psychological Measurement*, 78(1), 128-166.
- Ma'ruf, Ahmad A. (2007). *Analisis DIF Berbasis Jenis Kelamin Soal Olimpiade Biologi SLTP 2007 Tingkat Propinsi di DIY*. Jakarta: Universitas Indonesia.
- Prasetyo. (2010). Kelemahan dan Kekurangan Tes Obyektif. Online. <https://bocahpringsewu.wordpress.com/2010/11/01/kelemahan-dan-kekurangan-tes-objektif/> tanggal 18 Januari 2016 pukul 14.15 WIB.
- _____.(2014).”Teori Respon Butir dan Penerapannya”.Yogyakarta: Parama Publishing.
- Sudaryono.(2013). Implementasi Teori Responsi Butir (*Item Response Theory*) pada Penilaian Hasil Belajar Akhir di Sekolah. Online. <http://sudaryono.ilearning.me/2013/01/18/implementasi-teori-responsi-butir-item-response-theory-pada-penilaian-hasil-belajar-akhir-di-sekolah/> tanggal 24 Januari 2016 pukul 16.02 WIB.
- Sumintono, B & Whidhiarso, W. (2015) . *Aplikasi Permodelan Rasch Pada Assessment Pendidikan*. Trim Komunikata.
- Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning. *In 1st International Conference on Education Innovation (ICEI 2017)*. Atlantis Press.
- Susongko, Purwo.(2001). *Keberfungsian Butir Diferensial Perangkat Tes Ebtanas Kimia Sekolah Menengah Umum di Jawa Tengah*. Jurnal Penelitian dan Evaluasi. Nomor 4 Tahun III.
- _____. (2013). ”Penilaian Hasil Belajar”.Tegal: Badan Penerbit Universitas Pancasakti Tegal.
- _____. (2014). ”Pengantar Metodologi Penelitian Pendidikan”. Tegal: Badan Penerbit Universitas Pancasakti Tegal.
- _____.(2019). *Aplikasi Model Rasch Dalam Pengukuran Pendidikan Berbasis Program R*. Tegal. Universitas Pancasakti Tegal.
- Retnawati. (2014). *Teori Respon Butir dan Penerapannya* .Yogyakarta: Parama Publishing.
- Suwarto. (2011). Teori Tes Klasik dan Teori Tes Modern. Online. <http://download.portalgaruda.org/article.php?article=268350&val=7107&title=Teori%20Tes%20Klasik%20dan%20Teori%20Tes%20Modern/> tanggal 18 Januari 2016 pukul 12.58 WIB
- _____. (2012). *Pengembangan Tes Diagnostik dalam Pembelajaran*. Yogyakarta: Pustaka Pelajar
- Triyanto, Nur.(2014).*Bias Gender Ujian Akhir Semester Genap Fisika Kelas X SMA Negeri Kabupaten Purworejo Tahun Pelajaran 2012/2013*. Skripsi Universitas Muhammadiyah Purworejo
- UPS.(2019).*Pedoman Penyusunan Skripsi*. Tegal: Badan Penerbit Universitas Pancasakti Tegal.
- Wenning, C,J.& Vierya,R.(2015).*Teaching High School Phycics.1*. Publisher: Washington DC: National Academies Press.
- Wenning, C,J. (2006). Assesing nature of science literacy as one component of scientific literacy. *Journal of Physical Teacher Education Online*,3(4), 3 – 14.
- _____.(2007). Assesing inquiry skill as a component of scientific literacy. *Journal of Physical Teacher Education Online*,3(4), 3 – 14.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language assessment quarterly*, 4(2), 223-233.