**PENDETEKSIAN BIAS BUTIR SOAL MATEMATIKA UJIAN SEKOLAH SMP NEGERI DI KABUPATEN BARRU BERDASARKAN KELOMPOK GENDER, STATUS SOSIAL, DAN WILAYAH**

*Detection Of Bias Mathematics Item Of School Final Examination At Smp Negeri In Barru By Group Gender, Social Status, And Regional*

**Nurhidayah**

**Program Pascasarjana UNM, PEP, Nurhidayah\_UIN@yahoo.com**

**ABSTRAK**

*Pentingnya pendeteksian bias dilakukan untuk mendapatkan soal yang benar-benar mengukur kemampuan siswa tanpa memihak pada kelompok tertentu sehingga menghasilkan soal yang berkualitas. Oleh karena itu, penelitian ini bertujuan untuk mendeteksi bias butir yang terdapat dalam soal matematika UAS pada kelompok gender, status sosial, dan wilayah berdasarkan teori respon butir. Jenis penelitian ini eksploratif, yaitu akan mengungkap butir yang terdeteksi bias dalam soal UAS berdasarkan teori respon butir. Subjek dalam penelitian ini adalah butir perangkat soal dengan responden seluruh lembar jawaban peserta tes di Kabupaten Barru dengan jumlah responden 469 siswa. Sebelum melakukan pendeteksian bias butir terlebih dahulu dilakukan estimasi parameter dengan mengunakan model logistik 2 parameter untuk melihat daya beda dan tingkat kesukaran butir pada perangkat soal. Selanjutnya, bias dideteksi dengan model logistik 2 parameter dengan mengunakan metode Lord’s chi-square berdasarkan teori respon butir. Dari hasil pengolahan data, butir soal matematika UAS SMP Negeri 1 Tanete Rilau terindikasi memuat bias gender, bias wilayah, dan bias status sosial.*

***Kata Kunci*** *: Teori Respon Butir, Bias Butir, Gender, Wilayah, Status Sosial*

**PENDAHULUAN**

Evaluasi menjadi kegiatan yang memegang peranan penting dalam mengukur keberhasilan proses pembelajaran. Salah satu bentuk pengukuran siswa terhadap hasil belajar adalah Ujian akhir sekolah atau UAS yang merupakan tes akhir siswa di Sekolah. Selain ujian akhir nasional, ujian akhir sekolah juga merupakan tes pengukuran yang penting dalam melihat hasil belajar siswa karena hasil ujian sekolah juga berfungsi dalam memberikan gambaran dari kemampuan siswa secara murni dan memperlihatkan keberhasilan proses pembelajaran secara keselurahan ditingkat sekolah tersebut.

Karena itu, Instrumen tes yang digunakan untuk mengukur seharusnya memiliki validitas dan reliabilitas yang mantap serta adil. Adil artinya butir-butir tes tidak menguntungkan atau merugikan sekelompok peserta tes tertentu. Tes yang menguntungkan atau merugikan kelompok tertentu dikatakan tes tersebut memuat bias butir. Artinya, tidak ada orang atau kelompok orang tertentu yang merasa dirugikan dengan adanya butir soal yang tidak adil itu. Kenyatannya tidak selalu demikian, soal ujian nasional masih saja ada yang mengandung bias butir seperti yang di paparkan Retnawati (2005) menyatakan bahwa perbedaan jenis kelamin menurut Osterlind (1983), Camili dan Shepard (1994), Hambelton dan Rogers (1995) serta Sceheunem dan Bleisten (1999) dapat menimbulkan bias butir.

Camili & Shepard (1994: 3) menyatakan “bias adalah konsep yang didefinisikan dalam hal kelompok peserta ujian dimana seorang individu yang mungkin memiliki kesulitan menunjukkan apa yang dia tahu tentang tes tersebut”.

Hal ini senada dengan pendapat Hambleton (dalam Ilyas, 2011: 115) yang menyatakan “soal yang mengandung bias adalah adanya perbedaan skor perolehan yang disebabkan oleh adanya unsur yang menguntungkan atau merugikan bagi peserta”.

Hal ini ini mengindikasikan bahwa kelompok gender, status sosial, dan wilayah rentan menimbulkan bias pada perangkat soal. Termasuk pada kelompok peserta didik yang terdapat pada wilayah kabupaten Barru.

“Bias tes yang paling sering muncul adalah pada perbedaan kelompok pada jenis kelamin, perbedaan ras dan etnis. Namun, bias tes juga dapat menjadi masalah bagi subkelompok lain seperti kelas sosial, usia, wilayah, dan sejenisnya”. (Camili & Shepar, 1994 : 3)

Dilihat dari wilayah Kabupaten Barru yang terdiri dari 37 SMP Negeri yang terbagi dari 7 Kecamatan dengan beraneka macam perbedaan baik perbedaan kehidupan sosial, ekonomi, dan budaya. Perbedaan tersebut menyebabkan latar belakang siswa yang mengikuti UAS di tiap Sekolah tidak sama. Keadaan demikian sangat potensial terdapat *DIF* pada item-item tes UAS yang disebabkan latar belakang siswa tiap-tiap sekolah yang berbeda.

Akan tetapi, pada kenyataanya masih banyak guru-guru di sekolah yang kurang menyadari pentingnya pendeteksian bias dilakukan untuk mendapatkan soal yang benar-benar mengukur kemampuan siswa dalam menghasilkan soal yang benar-benar berkualitas. Bahkan beberapa diantara mereka yang mengetahui tentang bias hanya cenderung membiarkan dan tidak peduli pentingnya pendeteksian bias guna menghasilkan soal yang benar-benar mengukur tingkat kemampuan siswa secara objektif.

Untuk itu dalam penelitian ini, pendeteksian bias pada butir tes perangkat soal matematika UAS SMP ditinjau berdasarkan pengelompokan siswa dari segi gender, wilayah, dan satus social. Dimana gender akan dibagi kelompok untuk laki-laki dan perempuan, wilayah untuk kelompok perkotaan dan pedesaan, dan status sosial untuk pengelompokan mampu dan kurang mampu yang selanjutnya akan disebut sebagai kelompok fokus dan kelompok pembandingnya.

Hal ini didukung oleh pendapat, Camili dan Separd.(1994:3) yang menyatakan bahwa Bias tes yang paling sering menjadi masalah dalam instrumen adalah perbedaan kelompok jenis kelamin dan perbedaan ras dan etnis. Namun, bias dalam tes juga dapat menjadi terjadi bagi subkelompok yang terdiri oleh kelas sosial, usia, menurut wilayah perkotaan atau pedesaan, dan sejenisnya.

Menurut Budiono (2005 : 171), “Suatu butir disebut baik apabila adil terhadap peserta tes yang mempunyai kemampuan yang sama walau berasal dari kelompok yang berbeda. Pengujian untuk melihat apakah butir soal bertindak adil atau tidak, disebut pengujian keberadaan fungsi butir *(differential item functioning, DIF)”.*

*DIF* dapat disebabkan oleh bermacam-macam hal, seperti perbedaan budaya, perbedaan pendidikan, perbedaan jenis kelamin dll. Sebuah item yang baik seharusnya berlaku adil bagi setiap peserta tes, tidak boleh menguntungkan suatu kelompok tertentu seperti menguntungkan kelompok perempuan karena item tes tentang suatu hal yang berhubungan erat / umumnya diketahui oleh kaum perempuan, sehingga merugikan kaum laki-laki.

Dalam bias butir dikenal dua macam bias, yaitu bias eksternal dan bias internal. Bias eksternal menurut Osterlin (dalam Ratnawati, 2014: 127) “merupakan suatu derajat pada sekor tes yang menunjukkan hubungan korelasioanl dari variable-variabel bebas suatu tes. Bias eksternal diantaranya adalah konsekuensi sosial dalam penggunaan tes seperti keadilan dalam penyelenggaraan tes dan kriteria seleksi yang digunakan, maksudnya adalah pihak penyelenggara tes berwenang menyelenggarakan tes dan membuat kriteria seleksi dengan adil. Dengan demikian yang menjadi perhatian dan titik berat dalam bias eksternal adalah keseluruhan tes baru pada butir-butir tes itu sendiri”.

Karena itu, faktor-faktor yang mempengaruhi *DIF* dibedakan atas faktor internal dan faktor eksternal. Faktor internal adalah faktor-faktor yang berhubungan dengan karakteristik butir soal, sedangkan faktor eksternal adalah faktor-faktor yang berhubungan dengan kondisi responden.

Untuk itu, lanjut menurut Sudaryono (2012 : 135-137) bias dapat dibagi menjadi dua macam yaitu *DIF uniform* (konsisten) dan *DIF non uniform* (tidak konsisten). *DIF uniform* muncul jika keuntungan salah satu kelompok terhadap kelompok lainnya terjadi pada setiap level kemampuan, sedangkan *DIF non uniform* muncul jika keuntungan salah satu kelompok terhadap kelompok lainnya tidak terjadi pada setiap level kemampuan. Hal ini menjelaskan bahwa untuk satu kelompok memiliki rentang kemampuan yang tidak sama. Hal ini terjadi ketika dua *ICCs* memiliki parameter daya beda yang sama.

**Metode**

Penelitian ini merupakan Jenis Penelitian exploratif dengan pendekatan deskriptif. Penelitian ini terdiri dari dua kategori utama, yaitu : (1) penelitian karakteristik butir pada perangkat tes, dan (2) penelitian identifikasi *DIF* pada butir-butir tes yang digunakan yang dilakukan berdasarkan teori respon butir. Penelitian ini dilakukan di SMP Negeri 2 Barru, SMP Negeri 1 Tanete Riaja, dan SMP Negeri 2 Tanete Riaja dimana kegiatan penelitian dilaksanakan pada bulan Maret-Mei 2016 dengan objek penelitian berupa soal UAS yang berasal dari SMP Negeri 1 Tanete Rilau dengan jumlah 40 butir soal. Subjek penelitian yaitu seluruh lembar jawaban yang menjadi respon dari peserta tes yang terdiri dari siswa kelas IX SMP Negeri 2 Barru, SMP Negeri 1 Tanete Riaja, dan SMP Negeri 2 Tanete Riaja tahun pelajaran 2015/2016. Pemilihan subjek dalam penelitian ini dilakukan dengan menggunakan teknik purposif sampling sebanyak 469 respon peserta didik.

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi, yaitu dengan megutip respon peserta Ujian Sekolah pada mata pelajaran Matematika di SMP Negeri di Kabupaten Barru tahun pelajaran 2015/2016.

Selanjutnya, sebelum melakukan pengolahan data instrumen yang digunakan dalam penelitian ini terlebih dahulu dilakukan validasi untuk menyakinkan bahwa butir soal yang digunakan valid maka diadakan validasi secara content oleh pakar. Hasil analisis validasi butir berdasarkan aspek pengetahuan (tes hasil belajar matematika) yang diawali dengan memberikan hasil rancangan instrument pada validator. Validasi isi oleh pakar melibatkan dua orang pakar, yakni Dr. Triyanto Pristiwaluyo, M.Pd dan Dr. Rusyadi, M.Pd, dimana Keduanya merupakan dosen PEP Pascasarjana UNM.

Dari penilaian yang diberikan oleh kedua validator dihitung tingkat validasinya berdasarkan rumus konsistensi internal model Gregory diperoleh hasil koefisien konsistensi internal sebesar 1 atau 100%*.* Sebagaimana yang diketahui bahwa suatu pengukuran dapat dinyatakan valid jika hasil penilaian dari kedua validator memiliki validitas isi > 0,75 (Ruslan, 2009: 18-19). Hal ini berarti bahwa hasil penilaian kedua validator memiliki “relevensi yang kuat”.

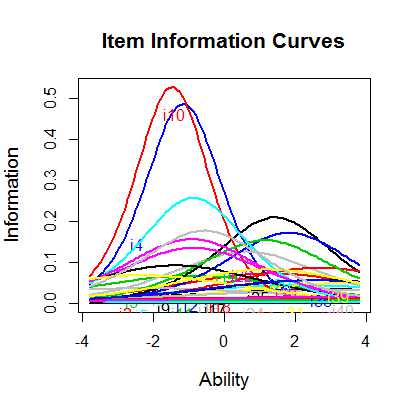
Selanjutnya, Analisis terhadap karakteristik butir tes UAS dilakukan melalui pendekatan teori respon butir dengan analisis data kuantitatif berdasarkan model logistik 2 parameter. Kualitas butir ditelaah dengan menghitung parameter butir meliputi (a) daya beda (b) tingkat kesukaran, dan (c) informasi butir.

**HASIL DAN PEMBAHASAN**

Pada penelitian ini, analisis butir karakteristik soal matematika Ujian Sekolah SMP Negeri 1 Tanete Rilau Tahun Pelajaran 2015/2016 dilakukan dengan menggunakan Program R versi 3.2.3 menggunakan model 2 parameter (2P), dimana karakteristik butir meliputi: (1) Daya beda, (2) Tingkat Kesukaran, dan untuk melihat keberfungsian butir dilakukan pula estimasi untuk(3) fungsi informasi.

Berdasarkan hasil pengolahan data yang telah dilakukan, karakteristik butir berdasarkan daya beda menunjukkan bahwa dari 40 butir soal yang dianalisis secara keseluruhan tanpa memisahkan kelompok perempuan dan kelompok laki-laki, diperoleh 29 atau 72.5% butir soal dengan kategori baik dan 11 butir soal atau 27.5% berkategori buruk dalam mengukur kemampuan peserta tes baik yang berkemampuan tinggi ataupun yang berkemampuan rendah.

Sedangkan karakteristik butir tingkat kesukaran atau disimpolkan *b* menunjukkan bahwa dari 40 butir soal yang diberikan kepada 469 peserta didik dilihat dari parameter tingkat kesukaran diperoleh bahwa sebanyak 38 butir atau 95 % dan 1 soal dengan kategori mudah dan 1 soal dengan kategori sangat mudah. Karna itu, tingkat kesukaran soal pada tes ini berada pada kategori sangat sukar sehingga butir soal dalam tes ini tidak dapat melakukan fungsi ukurnya dengan baik sehingga di asumsikan bahwa peserte tes cenderung menjawab dengan tebakan.

Selain dari daya pembeda butir soal dan tingkat kesukaran butir soal, dilakukan pula estimasi terhadap fungsi informasi, hal ini dimaksudkan untuk memberitahu penyusun tes mengenai seberapa baik kemampuan pada masing-masing level. Fungsi informasi tidak tergantung pada distribusi tes karena teknik ini merupakan aplikasi dari teori respon butir. Berdasarkan hasil dari program R versi 3.2.3 menurut teori respon butir, parameter fungsi informasi dapat dilihat pada gambar dibawah ini;

Gambar 4.5 Kurva Informasi Butir Tes Soal Ujian Akhir Sekolah Tahun Pelajaran 2015/2016 SMPN 1 Tanete Rilau berdasarkan Teori Respon Butir Model Logistik 2P

Berdasarkan gambar kurva, dapat terlihat bahwa informasi tiap butir berbeda dimana grafik tiap butir yang ditandai dengan warna yang berbeda. Titik puncak kurva yang berada dibawah kemapuan nol memberikan informasi bahwa butir soal cocok diberikan kepada peserta tes yang berkemampuan rendah dan puncak kurva yang berada diatas kemampuan nol memberikan informasi bahwa butir soal tersebut cocok diberikan kepada peserta tes dengan kemampuan tinggi.

Berdasarkan grafik kurva diatas menunjukkan kurva yang mempunyai informasi paling tinggi (puncak kurva) yaitu berada pada level kemampuan mendekati -2, sedangkan informasi paling rendah diberikan pada level kemampuan 4.

Selanjutnya dalam pendeteksian *DIF* pada kelompok gender, status sosial, dan wilayah terlebih dahulu ditentukan kelompok fokus dan kelompok referensi (pembandingnya). Hal ini sejalan dengan pendapat McNamara & Roever (dalam Karami & Nodoushan, 2011: 134) yang menyatakan “Setidaknya ada dua kelompok dalam penelitian *DIF*, yaitu kelompok fokus dan kelompok referensi. Kelompok fokus atau sekelompok minoritas adalah kelompok yang berpotensi dirugikan. Kelompok yang dianggap berpotensi diuntungkan dengan uji disebut kelompok referensi”

Untuk itu dalam penelitian ini kelompok fokus dan kelompok referensi pada gander dibagai atas laki-laki sebagai kelompok fokus dan perempuan sebagai kelompok pembanding (referensi). Pada kelompok wilayah dibagi atas kota sebagai kelompok fokus dan desa sebagai kelompok pembanding. Pada kelompk status sosial kelompok fokus dan acuannya dibagi atas mampu sebagai kelompok fokus, dan kurang mampu sebagai kelompok pembanding. Menurut (Bachman, 2004) “penamaan kelompok tidak selalu jelas. Artinya, pelabelan kelompok bisa sewenang-wenang dilakukan oleh peneliti yang dianggap kelompok yang berpotensi dirugikan dan diuntungkan”

Selanjutnya Bias yang dideteksi berdasarkan teori respon butir model 2 parameter dengan menggunakan metode *Lord’s chi-square* dan untuk melihat terjadi *DIF* konsisten (*uniform)* atau *DIF* tidak konsisten (*non uniform*) dengan cara melihat grafik nilai peluangnya Dengan tingkat signifikansi yang digunakan 2 < 0,05 untuk butir yang tedeteksi bias. Hasil pendeteksian bias gender dari 40 butir soal diperoleh 5 Butir yang terindikasi bias konsisten dan terdapat 15 butir yang terjangkit bias yang tidak konsisten. Pada bias bias lokasi (wilayah) terdeteksi *DIF* sebanyak 7 butir jenis *DIF* yang konsisten, serta terdapat 15 butir jenis *DIF* yang tidak konsisten. *DIF* berdasarkan kelompok wilayah, dilakukan pula pendeteksian *DIF* berdasarkan status sosial, dimana pada kelompok ini terindikasi lebih banyak *DIF* yaitu 26 butir. Pada bias status sosial di identifikasi *DIF* konsisten sebanyak 13 butir dan *DIF* yang tidak konsisten sebanyak 13 butir soal.

Dari hasil pendeteksian bias *(DIF*) diidentifikasi terdapat 13 butir soal yang sama terjangkit *DIF* gender, *DIF* lokasii (wilayah) dan *DIF* status sosial. Untuk itu butir tes Ujian Akhir Sekolah SMP Negeri 1 Tanete Rilau tahun ajaran 2015/2016 terdeteksi terjangkit bias tes.

Hal ini sejalan dengan pendapat Kamata & Vaughn (dalam Karami & Nodoushan, 2011:133) menyatakan bahwa Differential fungsi butir (DIF) terjadi setiap kali peserta tes dari tingkat kemampuan yang sama tetapi dari kelompok yang berbeda memiliki peluang menjawab butir soal dengan benar.

Menurut Lord (1980 : 122) mengemukakan bahwa “ jika pada suatu tes memuat butir-butir yang memihak pada kelompok tertentu, maka tes tersebut dikatakan memuat bias atau mengandung keberfungsian tes diferensial (Differential Tes Functioning)”.

Berdasarkan teori diatas maka disimpulkan bahwa tes ujian sekolah SMP Negeri 1 Tenete Rilau dengan tes yang diberikan di tiga sekolah lain diantaranya SMP Negeri 1 Tanete Riaja, SMP Negeri 2 Tanete Riaja, dan SMP Negeri 2 Barru secara otomatis mengandung unsur bias butir. Faktor lain yang menyebabkan suatu tes tidak adil pada salah satu kelompok peserta uji tes ada dua yaitu, 1) faktor internal yaitu faktor dari dalam butir itu sendiri. Artinya jika kajian butir difokuskan pada komponen butir misalnya bentuk butir, materi butir, kalimat dan kata-kata yang digunakan, gambar, petunjuk serta objek atau stimulus yang digunakan dalam butir soal, 2) faktor dari luar butir yang disebut dengan faktor eksternal. Faktor eksternal ini terjadi akibat dari beberapa penyebab yaitu penyebab dari peserta didik itu sendiri, dari pelaksana tes yang dalam hal ini adalah guru dan pengawas, dan dari lingkungan. Namun, faktor-faktor ini masih perlu penelitian yang lebih mendalam.

Faktor dari peserta didik itu sendiri biasanya diakibatkan karena kondisi psikologis dan kesehatan pada saat mengikuti tes atau karena kebiasaan penggunaan bahasa di lingkungan keluarga serta lingkungan sekitarnya.Faktor dari guru ini terjadi ketika anak didik dalam memaknai ucapan atau penjelasan dari guru yang kurang memahami materi yang disajikan oleh guru. Faktor dari pelaksana tes terjadi karena pengawas pada saat itu tidak jeli atau kurang mengawasi pada saat tes berlangsung sehingga peserta didik bebas menyontek pada buku catatan atau sesama peserta didik yang lain (peserta tes). Faktor lingkungan terjadi karena suasana lingkungan pada saat itu kurang mendukung misalnya sekitar ruangan ujian yang terlalu ribut oleh kelas lain yang tidak belajar sehingga mengganggu konsentrasi peserta tes.

**SIMPULAN**

Berdasarkan hasil analisis data dan pembahasan penelitian yang telah dikemukakan dan dihubungkan dengan rumusan masalah, ada beberapa hal yang berkaitan dengan deteksi bias butir (DIF) pada Soal Matematika Ujian Akhir Sekolah SMP Negeri 1 Tanete Rilau yaitu; Karakteristik soal matematika Ujian Akhir Sekolah di SMP Negeri 1 Tanete Rilau tahun Ajaran 2015/2016 berdasarkan kelompok gender, dari 40 butir soal 20 butir soal terjangkit bias butir dimana 6 butir soal yang menguntungkan kelompok laki-laki, dan 14 butir soal yang menguntungkan kelompok perempuan. Bias butirberdasarkan kelompok wilayah terdapat 22 butir soal yang terjangkit bias wilayah (lokasi) dimana 11 butir soal yang menguntungkan kelompok yang berasal dari kota dan 11 butir yang menguntungkan kelompok yang berasal dari desa dan bias berdasarkan kelompok status sosial, dari 40 butir soal terdapat 26 butir soal yang terjangkit bias status sosial dimana 15 butir soal yang menguntungkan kelompok yang berasal dari status sosial yang mampu dan terdapat 11 butir soal yang menguntungkan kelompok yang berasal dari status sosial yang kurang mampu.

**SARAN**

Berdasarkan hasil yang diperoleh dalam penelitian, terdapat beberapa hal yang dianggap perlu untuk diperhatikan terkait dengan deteksi bias butir berdasarkan teori respon butir model 2P pada metode Lord’s chi-square statistic diantaranya; kepada penyusun soal matematika SMP baik dalam skala lokal maupun skala nasional sebaiknya menggunakan soal yang secara empiris terbukti berkualitas baik dan tidak mengandung DIF, pembuktian secara empiris terhadap kualitas suatu butir perlu lebih dikemabangkan dan disosialisasikan pada semua praktisi pendidikan misalnya dengan mengadakan pelatihan bagi guru-guru baik bidang studi matematika maupun bidang studi lainnya, untuk mendeteksi DIF suatu tes dapat dilakukan berdasarkan teori respon butir, untuk peneliti lebih lanjut seharusnya menggunakan instrumen observasi sebagai pendukung hasil penelitian

**DAFTAR PUSTAKA**

Adedoyin. O. O . 2010. *Using IRT approach to detect gender biased items in public examinations: A case study from the Botswana junior certificate examination in Mathematics*. Botswana: University of Botswana, Educational Research and Reviews Vol. 5 (7).

Arikunto, Suharsimi., & Cepi, Safruddin. *Evaluasi Program Pendidikan*. Jakarta: Bumi Aksara. Cet ke 5. 2014

Azward, Saifuddin, 1998.  *Dasar-Dasar Psikometri*. Yogyakarta: Pustaka Pelajar.

Budiono. 2005. Perbandingan Metode Mantel-Heanzel. Sibtest, regresi Logistik, dan perbedaan Peluang dalam mendeteksi keberbedaan fungsi butir. Surakarta: Jurnal penelitian dan evaluasi pendidikan, nomor 2, tahun VII.

Camilli, G., & Shepard, L. A. 1994. *Methods fo Identifying Biased Test Items*. Thousand Oaks, CA: sage Publication.

Nutrisiana, Destian. 2013. *Pengaruh motivasi belajar, cara belajar, dan kemampuan sosial-ekonomi orang tua terhadap hasil belajar mata pelajaran ekonomi siswa kelas xi ips ma al-asror semarang tahun ajaran 2012/2013.* Semarang: Economic Education Analysis Journal.

Djaali. & Muljono. 2008. *Pengukuran dalam bidang pendidikan*. Jakarta: Grasindo.

Hadi, Samsul.  *Keberfungsian Item Differensial Pada Perangkat Tes Ujian Nasional Matematika Sekolah Menengah Atas Di Jawa Tengah*. Semarang: Seminar Nasional Evaluasi Pendidikan Tahun 2013, 479-486.

Hambleton, R. K., Swaminathan H., & Rogers, H. J., 1991. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publication Inc.

Ismail, Ilyas. M. 2011. *Differential Item Functional Keberbedaan Fungsi Butir*. Jakarta: Lentera pendidikan, vol 14 no. 1.

Kementrian Pendidikan dan Kebudayaan. 2005*. Peraturan Mentri Pendidikan dan Kebudayaan Nomor 5 Tahun 2005 tentang kriteria kelulusan peserta didik, penyelenggaraan ujian nasional, dan penyelenggaraan ujian sekolah/madrasah/pendidikan kesetaraan pada smp/mts atau yang sederajat dan sma/ma/smk atau yang sederajat*. Jakarta: Depertemen Pendidikan dan Kebudayaan.

Kusaeri., & Suprananto. 2012. *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu.

Karami*.*, *H., & Nodoushan., M., A. 2011. Differential Item Functioning (DIF): Current problems and future directions*. Iran: University of Tehran, International Journal of Language Studies (IJLS), Vol. 5(3)

Manoppo, Ian. 2012. *Bias pada Instrumen Tes Hasil Belajar.* (<http://ian.manoppo.blogspot.com>)

Mansyur., Harun ., & Suratno. 2009. *Asesmen Pembelajaran di Sekolah*. Yogyakarta: Multi Pressindo

Mansyur. 2009. *Pengembangan Model Assesmen For Learning pada Pembelajaran Matematika di SMP*. Yogyakarta: Disertasi PPs Universitas Yogyakarta.

Mardapi, D., 2008. *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendikia

Naga, Dali. S, 1992. *Pengantar Teori Sekor pada Pengukuran*. Pendidikan. Gunadarma. Jakarta

Peraturan Mentri Pendidikan Nasional. Nomor 20 Tahun 2007 tentang *Standar Penilaian Pendidikan dinyatakan bahwa salah satu prinsip penilaian adalah adanya unsur keadilan. Adil dalam arti bahwa penilaian tidak menguntungkan atau merugikan peserta didik karena berkebutuhan khusus serta perbedaan latar belakang agama, suku, budaya, adat istiadat, status sosial ekonomi, dan gender*. Jakarta: Depertemen Pendidikan dan Kebudayaan.

Purwanto. 2014. *Evaluasi hasil Pembelajaran*. Yogyakarta: Pustaka Pelajar.

Retnawati, Heri. 2014. *Teori Respon Butir dan penerapannya*. 2014. Yogyakarta: Nuha Medika.

Retnawati, Heri. 2005. *Keberfungsian Butir Diferensial pada Perangkat Tes Seleksi Masuk SLTP Mata Pelajaran Matematika*. *Jurnal Penelitian dan dan Evaluasi Pendidikan,(online).* Nomor 6 (staff.uny.ac.id, Diakses 15 Mei Tahun 2005).

Retnawati, Heri. 2005. *Mendeteksi Bias Butir tes matematika dan Ipa berdasarkan Gander.* Yogyakarta: Makalah seminar nasional penelitian, pendidikan, dan penerapan MIPA.

Retanawati, Heri., & Hidayati Kana. 2006. *Pendeteksian Bias Tes Dan Butir Perangkat Soal Matematika Ujian Nasional SLTP Berdasarkan Teori Respons-Butir*. UNY: Executive Summary

Ruslan. 2009. *Validitas Isi*. Buletin Pa’biritta No.10 Tahun VI, 18-19

Rohmah, Dwi, Yuniarti. 2013. *Pengaruh Sikap dan Gender Terhadap Prestasi Belajar Bahasa Indonesia pada Siswa SMP Negeri Kelas VII di Kecamatan Sleman Yogyakarta tahun 2013/2014*. Program Studi Pendidikan Bahasa dan Sastra Indonesia. Fakultas Bahasa dan Seni. Universitas Negeri Yogyakarta. Yogyakarta: skripsi.

Sudaryono. 2012. Kajian Metode Deteksi *differential item function* (*dif*) Butir Soal Ujian Nasional dengan Teori Tes Klasik. Tangerang*: Jurnal Pendidikan dan Kebudayaan,* Vol. 18, Nomor 2, Juni 2012. (<http://sippendidikan.kemendikbud.go.id>, Diakses 11 Desember 2014)

Wahyu., Sugeng., Den., i& Estu. 2014. *Pengukuran Indeks Etos Belajar Siswa Di Daerah Istimewa Yogyakarta*. Yogyakarta: Jurnal Kependidikan, Volume 44, Nomor 1.

Sugihastuti & Itsna. 2010. *Gender & Inferioritas Perempuan*. Yogyakarta.: Pustaka Pelajar, Cetakan ke2.

Tiro, Arif, Muhammad.2010. *Cara Efektif Belajar Matematika*. Makassar: Adira Publisher.

Widoyoko, Putro, Eko, 2014. Teknik Penyusunan Instrumen. Pustaka Pelajar. Yogyakarta.

Zumbo. B. D., 2007. *Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going*. Vancouver: The University of British Columbia, Department of ECPS, Language Assessment Quarterly, *4*(2), 223–233.