

**KEBERFUNGSIAN BUTIR DIFERENSIAL PERANGKAT TES UJIAN  
AKHIR SEKOLAH BERSTANDAR NASIONAL (UASBN)  
MATA PELAJARAN MATEMATIKA  
DI KOTA KENDARI**

*Oleh :  
Zulkarnain*

**Abstrak**

Penelitian ini bertujuan untuk mengungkap: (1) karakteristik butir perangkat tes yang digunakan pada UASBN menurut teori tes klasik dan teori respons butir, (2) kesalahan pengukuran menurut teori tes klasik dan teori respons butir, (3) butir-butir yang terdeteksi mengandung *DIF* berdasarkan perbedaan jenis kelamin (*gender*).

Analisis dalam penelitian ini didasarkan pada respons siswa peserta Ujian Akhir Sekolah Berstandar Nasional (UASBN) pada mata pelajaran Matematika SD di Kota Kendari, Sulawesi Tenggara tahun pelajaran 2007/2008 yang diikuti 122 SD dengan jumlah keseluruhan peserta 4.886 siswa. Sumber data berupa lembar jawaban komputer (LJK) yang diambil dengan teknik *purposive random sampling* sebanyak 1.259 yang terdiri dari 624 respons siswa laki-laki dan 625 respons siswa perempuan. Metode pendeteksian *DIF* berdasarkan perbedaan jenis kelamin (*gender*) yang digunakan dalam penelitian ini adalah metode Mantel-Haenszel.

Hasil analisis deskriptif berdasarkan teori tes klasik menunjukkan bahwa dari 40 butir soal yang dianalisis 28 butir yang berkualitas baik, indeks reliabilitas 0,810 dan Estimasi kesalahan pengukuran sebesar 2,779. Berdasarkan teori respons butir 3 parameter, 16 butir berkualitas baik dengan nilai informasi 7,728, Estimasi kesalahan pengukuran sebesar 0,359. Kesalahan pengukuran lebih besar jika menggunakan teori tes klasik. Analisis pendeteksian *DIF* dengan metode Mantel-Haenszel tidak ditemukan adanya butir yang terdeteksi mengandung *DIF* berdasarkan perbedaan jenis kelamin (*gender*)

*Kata kunci: Keberfungsian Butir Diferensial, UASBN dan Mantel Haenszel*

## ***Abstract***

This research aims at revealing: (1) the characteristics of test items based on the of classical test-theory and item-response theory, (2) the standard error measurement of classical test-theory and item-response theory, and (3) the items detected as containing Differential Item Functioning (*DIF*) based on gender.

Analysis of this study was based on testee's response to the Mathematics test in the Final Examination of National Standard at Elementary School in Kendari City, South-East Sulawesi in 2007/2008 academy year, consisting of 122 elementary schools with 4,886 testees. The data source is computer answer sheets of 1,259, 624 from male student responses and 625 female student responses, established using the *purposive random sampling* technique. This study employs the Mantel-Haenszel method, a detection method of *DIF* based on gender difference (gender).

The result of the descriptive analysis using the classical test-theory indicates that 25 out of 40 mathematics test items were *good*, with a reliability index of 0.810 and SEM of 2.779. Based on the item response theory 3 parameter, 16 items were *good*, with value information function was 7.728, and the SEM was 0.359. A higher measurement error of the test was found when using the classical test theory. The result of *DIF* analysis on the mathematics test employing Mantel-Haenszel method did not find any items containing *DIF* based on gender difference.

*Keyword: The Differential Item Functioning (DIF), Final Examination of National Standard School and Manthel-Haenszel*

## **PENDAHULUAN**

Petunjuk dan pelaksanaan Ujian Akhir Sekolah Berstandar Nasional (UASBN), oleh Mendiknas telah mengeluarkan Peraturan Menteri Pendidikan Nasional Nomor 39 Tahun 2007 tentang UASBN untuk Sekolah Dasar/Madrasah Ibtidaiyah/Sekolah Dasar Luar Biasa (SD/MI/SDLB) Tahun Pelajaran 2007/2008. UASBN adalah ujian nasional yang dilaksanakan secara terintegrasi dengan pelaksanaan ujian SD/MI/SDLB. Perubahan penyelenggaraan UASBN sebagai pengganti UASD pada dasarnya tidak ada perubahan yang mendasar, perbedaannya butir soal yang diujikan 25% dibuat pusat dan 75% dibuat oleh Dinas Provinsi (Depdiknas, 2007: 23). Karena porsi yang lebih besar yang dibuat oleh Dinas Provinsi, sehingga dipastikan UASBN tidak memberatkan peserta ujian.

Butir soal 25% buatan pusat adalah butir soal diambil dari bank soal yang dikembangkan Puspendik. Butir soalnya sudah dikalibrasi dan dideteksi tidak mengandung *DIF* atau bias butir. Untuk porsi 75% butir soal dibuat Diknas tingkat provinsi disusun oleh perwakilan guru dari setiap kabupaten/kota, butir soalnya belum diujicobakan dan dikalibrasi, sehingga bisa diduga mengandung bias butir (*DIF*). Dengan dasar itulah bias butir perangkat tes UASBN berpeluang mengandung *DIF* dan merupakan salah satu ketimpangan yang potensial terjadi.

Bias butir tes UASBN dapat dideteksi dengan berbagai teknik dan pendekatan untuk menentukan apakah butir-butir tes UASBN sesuai atau tidak dengan ketentuan psikometrik yang ada, yaitu berlaku untuk semua orang dalam populasi, tanpa

memperhatikan keanggotaan subkelompok yang ada dalam populasinya, seperti etnik/suku, jenis kelamin, tingkat sosial, usia, wilayah, lingkungan desa/kota, dan sejenisnya. Pendekatan pendeteksian dapat dilakukan dengan berdasarkan teori tes klasik dan pendekatan teori tes modern.

Pendeteksian *DIF* berdasarkan perbedaan jenis kelamin (*gender*) pada penelitian ini berdasarkan teori tes klasik. Pendeteksian *DIF* berdasarkan perbedaan jenis kelamin (*gender*) dilakukan dengan berbagai pertimbangan diantaranya (1) hanya atribut jenis kelamin yang dapat ditemukan pada dokumen peserta UASBN, (2) dari aspek geografis, peserta UASBN di Kota Kendari berada pada wilayah yang sama, (3) dari aspek usia, peserta UASBN memiliki rata-rata usia yang relatif sama, dan (4) adanya penelitian yang mendukung dan relevan dengan pendeteksian *DIF* berdasarkan jenis kelamin

Penelitian ini penting dilaksanakan dan bertujuan untuk: (1) mendeskripsikan karakteristik perangkat tes Ujian Akhir Sekolah Berstandar Nasional mata pelajaran matematika di Kota Kendari berdasarkan teori tes klasik dan teori respons butir model logistik tiga-parameter; (2) mengetahui besarnya kesalahan pengukuran perangkat tes Ujian Akhir Sekolah Berstandar Nasional mata pelajaran matematika di Kota Kendari berdasarkan teori tes klasik dan teori respons butir model logistik tiga-parameter; dan (3) mendeteksi seberapa banyak butir soal pada perangkat tes Ujian Akhir Sekolah Berstandar Nasional mata pelajaran matematika di Kota Kendari

memuat keberfungsian butir diferensial berdasarkan *gender* dengan menggunakan metode Mantel-Haenszel.

Matematika sebagai wahana pendidikan tidak hanya dapat digunakan untuk mencapai satu tujuan, misalnya mencerdaskan siswa, tetapi dapat pula membentuk kepribadian siswa serta mengembangkan keterampilan tertentu (Depdiknas, 2005a: 17). Pembentukan kepribadian siswa diarahkan melalui model dan pendekatan pembelajaran lewat nilai-nilai dalam kehidupan melalui matematika seperti jujur, disiplin, tepat waktu dan tanggung jawab. Untuk itu siswa perlu memiliki kemampuan memperoleh, memilih dan mengelola informasi dan pengetahuan yang aplikatif, realistik sebagai bekal untuk bertahan pada keadaan yang selalu berubah, tidak pasti dan kompetitif (Depdiknas, 2005b: 18).

Ciri utama matematika adalah penalaran deduktif, yaitu kebenaran suatu konsep atau pernyataan diperoleh sebagai akibat logis dari kebenaran sebelumnya, sehingga kaitan antar konsep atau pernyataan dalam matematika bersifat konsisten. Namun pembelajaran dan pemahaman konsep dapat diawali secara induktif melalui pengalaman peristiwa nyata atau intuisi. Proses induktif-deduktif dapat juga bersama-sama digunakan untuk mempelajari konsep matematika. Penerapan cara kerja matematika diharapkan dapat membentuk sikap kritis, kreatif, jujur, dan komunikatif (Depdiknas, 2005a: 21).

Salah satu instrumen yang digunakan dalam pengukuran prestasi belajar pada jenjang pendidikan adalah tes prestasi belajar. Menurut Djemari Mardapi (2008: 67),

tes merupakan sejumlah pertanyaan yang memiliki jawaban yang benar atau salah. Tes juga diartikan sebagai sejumlah pertanyaan yang membutuhkan jawaban atau tanggapan dengan tujuan mengukur tingkat kemampuan seseorang dari aspek tertentu. Hasil tes biasanya digunakan untuk memantau perkembangan mutu pendidikan. Tes yang digunakan untuk memantau mutu pendidikan harus baik dan berkualitas, yaitu memiliki kesalahan pengukuran yang sekecil mungkin.

Dalam proses pengukuran hasil belajar sangat diperlukan tes yang bermutu baik, karena baik-buruknya mutu tes akan menentukan mutu data yang dihasilkan. Tes prestasi belajar juga digunakan untuk mengukur sampel perilaku dan kemampuan seseorang yang menunjukkan tingkat pencapaian materi yang telah diajarkan. Idealnya, sesuai dengan taksonomi tujuan pendidikan, setiap evaluasi hasil belajar, harus mencakup baik ranah kognitif, afektif, maupun psikomotor. Namun kenyataannya, kebanyakan tes prestasi belajar hanya ranah kognitif saja yang ditonjolkan.

Ada dua tipe tes, termasuk tes belajar matematika yakni tes objektif dan tes uraian (*essay*). Tes objektif adalah tes yang telah disediakan pilihan jawabannya. Sedangkan tes uraian berupa soal yang masing-masing mengandung permasalahan dan menuntut penguraian sebagai jawaban. Ciri-ciri tes objektif diantaranya: (1) siswa bekerja terhadap tugas-tugas yang sudah distruktur secara sempurna; (2) siswa mencari jawaban dari pilihan yang telah disediakan; (3) mencakup materi/bahan yang cukup luas; dan (4) tiap soal dilengkapi dengan kunci. Ciri-ciri pokok tes uraian

(*essay*) adalah: (1) siswa mengorganisasikan sendiri jawaban; (2) jawaban berdasarkan pada kata-kata dan tulisan sendiri; (3) tes uraian terbatas pada sejumlah kecil pertanyaan saja; dan (4) penskoran yang subjektif, kekurangtelitian penilai dalam memeriksa jawaban menyebabkan penilai menilai kemampuan siswa secara subjektif (Depdiknas, 2005b: 17-20)

Prosedur Operasional Standar pelaksanaan Ujian Akhir Sekolah Berstandar Nasional (UASBN) tahun 2008, tertuang dalam surat keputusan Badan Standar Nasional Pendidikan Nomor 983/BSNP/XI/2007 tanggal 20 November 2007 tentang Prosedur Operasional Standar (POS) Ujian Akhir Sekolah Berstandar Nasional (UASBN) untuk Sekolah Dasar, Madrasah Ibtidayah, dan Sekolah Luar Biasa (SD/MI/SDLB) Tahun Pelajaran 2007/2008. Mengingat Peraturan Menteri Pendidikan Nasional Nomor 39 Tahun 2007 tentang Ujian Akhir Sekolah Berstandar Nasional (UASBN) SD/MI/SDLB tahun pelajaran 2007/2008 pada pasal 6, pasal 7, pasal 8 dan pasal 9 dijelaskan, mata pelajaran yang diujikan pada penyelenggaraan UASBN tahun pelajaran 2007/2008 meliputi Bahasa Indonesia, Matematika, dan Ilmu Pengetahuan Alam (IPA). Standar Kompetensi Lulusan UASBN (SKLUASBN) yang digunakan ketiga mata pelajaran yang diujikan merupakan irisan (*interseksi*) dari pokok bahasan/subpokok bahasan kurikulum 1994, Standar kompetensi dan kompetensi dasar kurikulum 2004, serta Standar Kompetensi dan Kompetensi Dasar pada Standar Isi sesuai pada Permen Nomor 22 tahun 2006. Penyelenggara UASBN adalah BSNP bekerja sama dengan Instansi terkait di lingkup pemerintah Pusat,

Pemda Provinsi, Pemda kabupaten/kota, serta Satuan Pendidikan. Untuk spesifikasi soal UASBN tahun pelajaran 2007/2008 ditetapkan oleh BSNP terdiri dari 25% paket soal dibuat oleh pusat dan berlaku secara nasional dan 75% soal yang dibuat oleh daerah penyelenggara UASBN (Depdiknas, 2007: 2-3).

Teori tes klasik disebut demikian, karena teori ini sudah lama dikembangkan dan diaplikasikan, namun tetap bertahan dan masih digunakan sekarang ini untuk keperluan yang terbatas, teori tes klasik memuat berbagai keunggulan dan kelemahan. Keunggulan teori tes klasik menurut Budiyo (2005: 30), antara lain: (a) menggunakan konsep sederhana untuk menentukan kemampuan peserta; (b) menggunakan konsep sederhana dalam menghitung koefisien validitas dan realibilitas tes serta menghitung nilai parameter butir soal; (c) dapat digunakan pada sampel kecil, misalnya pada tingkat kelas; dan (d) sudah digunakan dalam kurun waktu yang lama pada praksis pengukuran dan pengujian, sehingga telah diketahui dan dipercaya oleh sebagian besar orang yang berkecimpung atau terlibat dengan dunia pendidikan dan *psychology*.

Selanjutnya menurut Allen & Yen, (1979: 60)

Classical true-score theory involves an additive model. An observed test score  $X$  is the sum of two components: a stable true score  $T$  and a random error score  $E$ . Error scores on a test are assumed to be uncorrelated with true score on that test and with true and error scores on all others test.

Teori tes klasik yang juga disebut teori skor murni didasarkan pada model aditif, yaitu skor amatan merupakan penjumlahan dari skor sebenarnya dan skor



kesalahan pengukuran. Kesalahan pengukuran disini maksudnya adalah kesalahan yang tidak sistematis atau acak. Kesalahan ini merupakan penyimpangan secara

Teori pengukuran klasik terdapat keterbatasan karena bersifat *group dependent* dan *Item dependent* (Hambleton, Swaminathan & Rogers, 1991: 2-5). *Group dependent* artinya hasil pengukuran tergantung dari kelompok peserta yang mengerjakan tes. Jika tes diujikan kepada kelompok peserta dengan kemampuan tinggi, tingkat kesulitan butir soal akan rendah. Sebaliknya jika tes diujikan kepada peserta yang kemampuan rendah, tingkat kesulitan butir soal akan tinggi. *Item dependent* artinya hasil pengukuran tergantung dari tes mana yang diujikan. Jika tes yang diujikan mempunyai tingkat kesulitan yang tinggi, estimasi kemampuan peserta akan rendah. Sebaliknya jika tes yang diujikan tingkat kesulitannya rendah maka estimasi kemampuan peserta akan tinggi. Kelemahan pengukuran semacam ini tidak terdapat pada teori respons butir (*Item respons theory*).

Untuk mengatasi kelemahan-kelemahan yang ada pada teori tes klasik, maka para pakar pengukuran berupaya mencari model alternatif. Model alternatif yang diinginkan adalah model yang memiliki sifat antara lain: (a) karakteristik butir soal tidak tergantung pada kelompok peserta yang diberikan butir soal tersebut; (b) skor yang menyatakan kemampuan peserta tes tidak tergantung pada tes; (c) model tersebut dapat dinyatakan dalam tingkatan (*level*) butir soal tidak dalam tingkatan tes; (d) model tidak memerlukan tes paralel untuk menghitung koefisien realibilitas; dan

(e) model menyediakan ukuran yang tepat untuk setiap skor kemampuan (Hambleton, Swaminathan & Rogers, 1991: 5).

Selanjutnya Hambleton, Swaminathan & Rogers (1991: 7), menyatakan konsep yang mendasar dari teori respons butir adalah pertama, respons seorang subjek pada suatu butir dapat memprediksi faktor yang disebut *trait* (kemampuan) yang dimiliki. Kedua, hubungan antara kemampuan subjek pada suatu butir dan kemampuan yang mendasarinya dapat digambarkan oleh suatu fungsi menaik secara monoton yang disebut kurva karakteristik (*item characteristic curve*). Probabilitas respons terhadap butir oleh individu untuk menentukan karakteristik dari individu dan butir dapat digambarkan dalam fungsi matematis merupakan konsep yang paling mendasar dari teori respons butir.

Respons yang diberikan individu terhadap butir soal dapat digunakan untuk mengestimasi kemampuan seseorang dan butir melalui analisa statistik. Untuk mengestimasi kemampuan dibutuhkan parameter. Parameter yang digunakan adalah parameter yang membentang sepanjang garis kontinum variabel yang diukur, garis kontinum bergerak dari sebelah kiri yang menunjukkan kemampuan kurang, menuju sebelah kanan yang menunjukkan kemampuan tinggi, untuk mengestimasi butir diperlukan parameter butir. Parameter butir terdiri dari  $a_i$ ,  $b_i$  dan  $c_i$  dimana  $a_i$  sebagai indeks daya beda,  $b_i$  sebagai indeks tingkat kesukaran dan  $c_i$  adalah *pseudo guessing* (Naga, 1992: 221).

## **METODE PENELITIAN**

Penelitian ini bertujuan untuk mendiskripsikan karakteristi butir soal dan keberfungsian butir diferensial (*DIF*) perangkat tes UASBN dan digolongkan dalam jenis penelitian *survey*, dengan pendekatan deskriptif kuantitatif. Penelitian ini dilaksanakan selama tiga bulan, yaitu bulan Mei sampai dengan bulan Juli 2008 dan bertempat di Kantor Dinas Diknas Kota Kendari

Populasi dalam penelitian ini adalah seluruh lembar jawaban peserta ujian SD/MI se-Kota Kendari tahun pelajaran 2007/2008 mata pelajaran matematika. Sampel penelitian adalah lembar jawaban komputer (LJK) siswa peserta UASBN tahun 2008 mata pelajaran matematika yang telah didokumentasikan, pengambilan sampel dilakukan dengan teknik *purposive random sampling*. Prosedur penentuan lokasi pengambilan data penelitian dilakukan melalui dua tahap yaitu: (1) menentukan wilayah sumber data. Teknik pengambilan dilakukan secara *purposive* berdasarkan pertimbangan peneliti, (2) menentukan sumber data yang berasal dari setiap kecamatan pengambilan dilakukan secara acak (*random sampling*), setiap kecamatan diambil secara acak minimal 120 respons peserta UASBN secara berimbang dari dua kelompok yaitu respons siswa laki-laki dan respons siswa perempuan. Banyaknya data dalam penelitian ini adalah 1.259 respons siswa.

Kegiatan analisis data dalam penelitian ini meliputi dua pokok kegiatan, yaitu: (1) estimasi parameter butir dengan menggunakan teori tes klasik dan teori respons butir 3 parameter, serta estimasi parameter kemampuan dari masing-masing peserta

UASBN; dan (2) pendeteksian *DIF* dengan menggunakan metode Mantel-Haenszel. Prosedur estimasi parameter butir menggunakan dua pendekatan, yaitu pendekatan teori tes klasik dan pendekatan teori respon butir.

Pendekatan pertama dilakukan dengan menentukan karakteristik butir perangkat tes yang meliputi parameter: tingkat kesukaran, daya beda, reliabilitas, dan kesalahan baku pengukuran. Karakteristik parameter berupa tingkat kesukaran, daya beda, kesalahan pengukuran, dan reliabilitas ini, dianalisis dengan menggunakan program *ITEMAN Version 3,00 MicroCAT (tm) Testing* sebagai fase satu. Pendekatan kedua menggunakan teori respon butir model logistik tiga parameter. Pendekatan ini dilakukan untuk menaksir kemampuan siswa dan parameter butir serta nilai informasi butir dan nilai informasi tes. Untuk keperluan tersebut peneliti menggunakan program *Bilog\_MG v3. 0 (REV 19990104.13000)*, sebagai fase kedua dan ketiga.

Berdasarkan teori tes klasik, secara empiris kualitas butir ditentukan oleh statistik butir yang meliputi parameter tingkat kesukaran  $p_i$ , daya pembeda  $d_i$ . Berdasarkan pendekatan teori respons butir, secara empiris kualitas butir ditentukan oleh statistik butir yang meliputi parameter daya pembeda  $a_i$  (*slope*), tingkat kesukaran  $b_i$  (*threshold*), dan tebakan semu  $c_i$  (*asymptote*).

Pendeteksian ada atau tidaknya *DIF* untuk perangkat soal UASBN diawali dengan mengelompokkan respons/jawaban siswa berdasarkan jenis kelamin, selanjutnya dilakukan analisis keberadaan *DIF* dengan metode Mantel-Haenszel. Adapun prosedur pendeteksian *DIF* metode Mantel-Haenszel sebagai berikut: (1)

hasil estimasi kemampuan (*ability*) dari dua kelompok *R* (kelompok laki-laki) dan *F* (kelompok perempuan) pada fase 3 *Bilog*, selanjutnya dilakukan *matching* pada (*ability*) yang sama; (2) hasil *matching* kelompok *R* dan *F*, kemudian dibagi menjadi 3 sub kelompok yaitu, sub kelompok atas, sub kelompok tengah, dan sub kelompok bawah. Kedua kelompok tersebut diambil dari populasi yang sama dan mengerjakan tes yang sama pula (*homogen*), kemudian dari tiap kelompok itu dibagi menjadi tiga subkelompok. Kedua kelompok yang diambil diusahakan pengelompokannya setara; (3) menghitung nilai proporsi sebagai dasar dalam pendeteksian *DIF*, yaitu proporsi jumlah peserta yang menjawab benar dari kelompok (laki-laki dan perempuan) dari ketiga subkelompok; (4) mengubah respons dari huruf ke dalam bentuk angka yakni 1 untuk jawaban benar dan 0 untuk jawaban salah, bagi peserta yang memiliki kemampuan sama dari kedua kelompok; (5) perbandingan dilakukan pada subkelompok hasil pembagian yang setara dan dilakukan perbutir dari perangkat tes yang dianalisis, kemudian dibuat tabel kontingensi; (6) menghitung statistik menggunakan persamaan Mantel-Haenszel; (7) menentukan kriteria pengujian menggunakan statistik *Chi-kuadrat* untuk menguji signifikansi hipotesis nol; dan (8) menarik kesimpulan.

## **HASIL PENELITIAN DAN PEMBAHASAN**

### **1. Karakteristik Parameter Butir**

#### **a. Berdasarkan Teori Tes Klasik**

Berdasarkan hasil analisis pendekatan teori tes klasik yang dilakukan dengan program *Iteman Version 3,00 MicroCAT (tm) Testing* pada langkah pertama menunjukkan bahwa dari 40 butir yang dianalisis terdapat 28 butir tes yang baik atau 70% dan 12 butir soal yang tidak baik atau 30%. Butir soal yang memenuhi kriteria tersebut adalah butir soal nomor 2, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, dan 35. Butir soal yang tidak memenuhi kriteria tersebut adalah butir soal nomor 1, 3, 9, 16, 21, 24, 29, 34, 36, 37, 38, 39 dan 40.

Jika butir soal dilihat dari sumbernya, terdapat 7 butir soal yang baik dari 10 butir soal yang dibuat oleh pusat (Puspendik) atau 70% yaitu soal nomor 4, 8, 14, 20, 23, 26, dan 35. Sedangkan butir soal yang tidak baik adalah 3 butir atau 40%, yaitu butir nomor 3, 38 dan 39.

Hasil analisis menggunakan program *Iteman Version 3,00 MicroCAT (tm) Testing* perangkat tes UASBN secara keseluruhan berdasarkan tingkat kesukaran tes, jumlah butir soal yang berkategori sukar sebanyak 2 butir soal, berkategori sedang 27 butir soal, dan yang berkategori mudah 11 butir soal. Daya pembeda soal, jumlah butir soal yang berkategori sangat baik sebanyak 25 butir, berkategori baik sebanyak 11 butir soal, berkategori cukup baik sebanyak 3, dan tidak baik 1 butir soal.

Berdasarkan kriteria penerimaan butir soal yang baik menurut teori tes klasik adalah perangkat tes yang memiliki parameter tingkat kesukaran butir soal antara 0,3 – 0,7 dan indeks daya beda lebih besar atau sama dengan 0,20. Dengan demikian, maka ada 28 butir soal yang berkategori baik dan 12 butir soal yang berkategori tidak baik. Nilai indeks reliabilitas sebesar 0,810 diperoleh menggunakan program *Iteman Version 3,00 MicroCAT (tm) Testing*.

#### **b. Berdasarkan Teori Respons Butir**

Berdasarkan hasil analisis perangkat soal mata pelajaran matematika menggunakan program *Bilog\_MG v3. 0 (REV 19990104.13000)* fase kedua. Dari 40 butir yang dianalisis terdapat 32 butir soal yang baik atau 80% dan 8 butir soal yang tidak baik atau 20%. Butir soal yang memenuhi kriteria teori respons butir tersebut adalah butir soal nomor 2, 3, 4, 5, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 27, 29, 30, 31, 32, 33, 34, 35, 37, 38 dan 40. Butir soal yang tidak memenuhi kriteria tersebut adalah butir soal nomor 1, 6, 7, 24, 26, 28, 36 dan 39.

Butir soal yang memenuhi kriteria teori respons butir 3 parameter dan cocok model adalah butir soal nomor 4, 8, 10, 12, 13, 14, 18, 19, 21, 22, 23, 25, 30, 31, 38 dan 40. Butir soal yang tidak memenuhi kriteria teori respons butir 3 parameter dan tidak cocok model adalah butir soal nomor 1, 9, 24, 26 dan 35. Butir soal yang berkriteria baik buatan pusat (Puspendik) menurut teori respons butir, dari 10 butir

ada 8 butir atau 80% yang baik, yaitu butir nomor 3, 4, 8, 14, 20, 23, 29 dan 35. Sedangkan butir yang kurang baik ada 2 atau 20% yaitu butir nomor 26 dan 39.

Berdasarkan hasil analisis perangkat tes UASBN secara keseluruhan ditinjau dari tingkat kesukaran soal, maka ada 5 butir soal yang berkategori tidak baik dan 35 butir soal yang berkategori baik. Daya pembeda soal ada 39 butir soal yang berkategori baik dan 1 butir soal yang berkategori tidak baik. Tebakan semu ada 38 butir soal yang berkategori baik dan 2 butir soal yang berkategori tidak baik. Berdasarkan kriteria penerimaan butir soal yang baik berdasarkan teori respons butir, maka ada 32 butir soal yang berkategori baik dan sebanyak 8 butir soal yang berkategori tidak baik

Berdasarkan kecocokan butir dengan model, menunjukkan bahwa dari 40 yang dapat dianalisis terdapat 21 butir yang cocok dengan model dan 19 butir yang tidak cocok dengan model. Butir yang cocok dengan model terdiri dari 16 butir dengan kualitas baik dan 5 butir dengan kualitas kurang baik. Nilai fungsi informasi tes adalah 7,728 dan nilai ini tercapai pada kemampuan siswa ( $\theta$ ) sebesar 0,6250 .

## **2. Kesalahan Pengukuran Perangkat Tes UASBN**

Kesalahan pengukuran digunakan untuk memahami kesalahan yang bersifat *random* atau acak yang mempengaruhi skor peserta. Hasil skor yang diperoleh bila peserta ujian ke- $i$ , memperoleh skor  $X_i$ , dengan menggunakan pendekatan teori tes klasik diperoleh koefisien kesalahan pengukurannya 2,779 dan skor total maksimum



yang dapat dicapai adalah 40, sehingga skor sebenarnya yang dapat diperoleh peserta UASBN tersebut berkisar pada  $X_i \pm 2,779$ . Bila perhitungan kesalahan pengukuran menggunakan teori respons butir, diperoleh koefisien kesalahan pengukuran sebesar 0,359, dan skor maksimum yang dapat dicapai oleh peserta adalah 40, maka skor peserta ujian UASBN sebenarnya berkisar pada  $X_i \pm 0,359$ .

Berdasarkan hasil analisis dengan perhitungan menggunakan pendekatan teori tes klasik dan teori respons butir diperoleh kesimpulan, bahwa kesalahan pengukuran yang terkecil bila menggunakan pendekatan teori respons butir, dengan demikian pendekatan teori respons butir lebih baik, akurat teliti dan cocok untuk mengestimasi kesalahan pengukuran pada perangkat tes mata pelajaran matematika UASBN Kota Kendari tahun pelajaran 2007/2008.

### **3. Pendeteksian Keberfungsian Butir Diferensial (*DIF*)**

Hasil analisis menunjukkan bahwa dengan menggunakan metode Mantel-Haenszel tidak ditemukan adanya butir yang terdeteksi mengandung *DIF* berdasarkan perbedaan jenis kelamin pada perangkat tes matematika UASBN Kota Kendari tahun pelajaran 2007/2008. Tidak terdeteksi *DIF* pada perbedaan *gender* perangkat tes UASBN dimungkinkan karena rata-rata usia peserta UASBN cenderung hampir sama, sehingga belum ada perbedaan secara psikologi. Dengan demikian secara umum perangkat tes matematika USBN Kota Kendari ini termasuk kategori

perangkat tes yang baik, tidak diskriminasi, dipercaya, berlaku adil dalam mengukur apa yang seharusnya diukur.

## **KESIMPULAN**

Berdasarkan hasil penelitian perangkat tes UASBN SD/MI mata pelajaran matematika Kota Kendari tahun pelajaran 2007/2008 dapat disimpulkan sebagai berikut:

1. Karakteristik perangkat tes berdasarkan teori tes klasik, parameter tingkat kesukaran, ada 2 butir sukar, 27 butir sedang, dan 11 butir mudah. Jika ditinjau dari parameter daya pembedanya, maka ada 25 butir sangat baik, 11 butir baik, 3 butir cukup baik dan 1 butir tidak baik. Dengan demikian berdasarkan teori tes klasik, ada 28 butir yang baik, dan dapat digunakan untuk pengembangan bank soal. Perangkat tes tersebut mempunyai indeks reliabilitas sebesar 0,810
2. Karakteristik perangkat tes berdasarkan teori respons butir, parameter tingkat kesukaran (*threshold*), ada 35 butir baik dan 5 butir tidak baik. Jika ditinjau parameter daya pembedanya (*slope*), maka ada 39 butir soal baik dan 1 butir tidak baik. Jika ditinjau parameter tebakan semu (*pseudo guessing*) maka ada 38 butir baik dan, 2 butir tidak baik. Jika ditinjau kecocokan model 3 parameter, maka ada 21 butir cocok model, dan ada 16 butir dapat digunakan untuk pengembangan bank soal. Nilai fungsi informasi perangkat tes sebesar 7,728 dan nilai ini tercapai jika parameter kemampuan (*theta*) peserta tes sebesar 0,6250.

3. Kesalahan pengukuran (*Standard error of measurement*) berdasarkan teori tes klasik sebesar 2,779 dan kesalahan pengukuran berdasarkan teori respon butir sebesar 0,359. Teori respons butir lebih baik dari pada teori tes klasik dilihat dari estimasi kesalahan pengukurannya.
4. Tidak satu butirpun memuat keberfungsian butir diferensial (*DIF*) berdasarkan perbedaan jenis kelamin (*gender*) menggunakan metode Mantel-Haenszel.

#### **DAFTAR PUSTAKA**

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole Publishing Company
- Algina, J & Crocker, L (1986) *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston
- Camilli, Gregory & Shepard, Lorrie A. (1994). *Methods for identifying biased test items*. London: Sage publication.
- Depdiknas. (2007). *Permen Diknas Nomor 39 Tahun 2007 tentang Ujian Akhir Sekolah Berstandar Nasional (UASBN) untuk Sekolah Dasar/Madrasah Ibtidaiyah/Sekolah Dasar Luar Biasa (SD/MI/SDLB) Tahun Pelajaran 2007/2008*.
- \_\_\_\_\_. (2005a). *Wawasan pendidikan matematika*. Bahan pelatihan terintegrasi berbasis kompetensi guru SMP. Jakarta: Direktorat Pendidikan Lanjutan Pertama
- \_\_\_\_\_. (2005b). *Hakekat penilaian matematika*. Bahan pelatihan terintegrasi berbasis kompetensi guru SMP. Jakarta: Direktorat Pendidikan Lanjutan Pertama
- Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendikia Press Yogyakarta

- Gronlund, N. E. (1976). *Measurement and evaluation in teaching*. (6<sup>th</sup> Ed). New York: Macmillan Publishing Co., Inc
- Hambleton, R. K, Swaminathan, H & Rogers, J. H. (1991). *Fundamentals of item response theory*. London: Sage publications.
- Hambleton, R. K, & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer-Nijhoff, Publisher.
- Hulin, Drasgow & Parsons. (1983) *Item response theory: Application to psychological measurement*. Homewood: Dow Jones-Irwin.
- Holland, P. W., & Thayer, D.T. (1988). *Differential item functioning and the Mantel-Hansel procedure*. New Jersey: Lawrence Erlbaum.
- Lord, M. F. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum associates, publishers
- Mehrens, W. A & Lehman, I. J. (1973). *Measurement and evaluation in education and psychology*. New York: Hold, Rinehart and Winston, Inc.
- Osterlind, Steven, J. (1983). *Test item bias*. California: Sage Publication.
- Thorndike, R.M. (2005). *Measurement and evaluation in psychology and education* (7<sup>th</sup> Ed). New Jersey: Pearson Education Ltd.

### **Biodata**

Nama, **Zulkarnain, S.Pd**. Tempat dan tanggal lahir, **Ujung Pandang 13 November 1970**. Pendidikan terakhir, **Sarjana Pendidikan Matematika**. Bekerja sebagai **Guru SMP Negeri 8 Kendari**, Kota Kendari, Provinsi Sulawesi Tenggara.