

EXECUTIVE SUMMARY
DOSEN MUDA



PENDETEKSIAN BIAS TES DAN BUTIR
PERANGKAT SOAL MATEMATIKA UJIAN NASIONAL SLTP
BERDASARKAN TEORI RESPON-BUTIR

Oleh :

Heri Retnawati, M.Pd.

Kana Hidayati, M.Pd.

Dibiayai oleh
Direktur Jendral Pendidikan Tinggi Departemen Pendidikan Nasional
Sesuai Surat Perjanjian Pelaksanaan Penelitian
Nomor : 018/SP3/PP/DP2M/II/2006, tanggal 1 Februari 2006

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
JURUSAN PENDIDIKAN MATEMATIKA
UNIVERSITAS NEGERI YOGYAKARTA
September, 2006

EXECUTIVE SUMMARY
Pendeteksian Bias Tes dan Butir Perangkat Soal Matematika
Ujian Nasional SLTP Berdasarkan Teori Respons-Butir

Oleh : Heri Retnawati dan Kana Hidayati
Pendidikan Matematika FMIPA UNY

ABSTRACT

This research is aimed at revealing : the characteristics of the Junior High Schools nasional mathematics examination based on the differential test functioning (DIF) and the differential item functioning (DTF) occurred among the testees based on gender in DI Yogyakarta in 2004/2005.

The object of this study was the national mathematics examination tests at the JHSs in Yogyakarta city in 2004/2005 academic year. This study used tests and testees' responses to the test. The method was used to detect the differential item functioning of these tests is Likelihood Ratio test using BILOG 3.07 computer program.

The result indicated that using the classical test theory, allt items (30 items) of the Junior High Schools nasional mathematics examination were good items. The Junior High Schools nasional mathematics examination contained differential test functioning (DTF) favoured females a little at all ability. Based on item response theory, the test had 20 good items, 1 enough, 9 unfit items. Its value of information function was 9,469356 at 0,35 ability scale and its SEM coefficient was 9,469356. In the differential item functioning analysis on the test, there were 6 items containing DIF (item number 2, 8, 11, 15, 18, dan 25).

Keyword : item response theory, DTF, DIF

A. Pendahuluan

Ujian Nasional sebagai salah satu proses pengukuran terhadap hasil belajar tingkat nasional memiliki tujuan dan fungsi yang sangat penting dalam bidang pendidikan. Hasil Ujian Nasional yang diberikan dalam bentuk NUN (Nilai Ujian Nasional) sangat diharapkan dapat digunakan untuk melihat gambaran kemampuan siswa secara murni serta mampu berfungsi sebagai indikator keberhasilan proses pendidikan. Mengingat pentingnya NUN tersebut, maka perangkat soal yang disusun harus benar-benar dapat mengukur apa yang seharusnya diukur, memberikan hasil pengukuran yang dapat dipercaya, dan mencerminkan kemampuan siswa yang sebenarnya. Artinya, perbedaan skor yang

diperoleh seorang siswa dengan siswa lainnya semata-mata hanya karena perbedaan kemampuan di antara mereka, bukan disebabkan oleh faktor lain, seperti adanya bias pada butir tes yang mereka kerjakan.

Perangkat soal Ujian Nasional akan menjadi sesat jika tidak objektif, tidak adil, dan memihak kelompok tertentu sehingga dua siswa dari kelompok yang berbeda dengan kemampuan yang sama mempunyai peluang menjawab benar yang berbeda. Perangkat soal tersebut menjadi tidak baik kualitasnya karena mengandung sifat bias pada butir-butirnya. Oleh karena itu, pendeteksian terhadap ada tidaknya sifat bias pada tes dan butir-butir tes Ujian Nasional sangat perlu untuk dilakukan agar ketidakadilan atau kerugian pada kelompok tertentu dapat dihindarkan dan kemampuan siswa terukur secara objektif. Jika pada suatu tes memuat butir-butir yang memihak kelompok tertentu, maka tes tersebut dikatakan memuat bias atau mengandung keberfungsian butir diferensial (*Differential Item Functioning, DIF*) (Adams, 1992: 178). Adanya butir bias ini mengakibatkan suatu tes bersifat diskriminatif atau memihak pada kelompok tertentu yang penyebabnya dapat ditinjau dari berbagai segi, misalnya ras (etnik), budaya, wilayah, jenis kelamin, dan lain-lain (Osterlind, 1983: 179). Dalam pengukuran pendidikan, istilah bias pada suatu butir dikenal sebagai *Differential Item Functioning (DIF)* atau keberfungsian butir diferensial. Berbagai teknik atau metode pendeteksian *DIF* telah banyak ditemukan dan digunakan. Selain keberfungsian butir diferensial, untuk mengindikasikan suatu tes adil bagi semua golongan atau tidak, dapat digunakan keberfungsian tes diferensial (*Differential Test Functioning, DTF*). Keberfungsian tes diferensial ini merupakan hal yang penting bagi suatu perangkat soal karena *DTF* dapat menentukan apakah suatu tes adil bagi semua kelompok peserta tes atau tidak.

Berdasarkan uraian di atas, penelitian ini akan mengungkap bias tes dan butir pada perangkat soal Matematika Ujian Nasional tahun pelajaran 2004/2005 di Propinsi Daerah Istimewa Yogyakarta. Dalam hal ini metode pendeteksian *DIF* dan *DTF* yang dipilih adalah yang berdasarkan teori respons-butir. Pembagian kelompok siswa dalam penelitian ini dilakukan dengan berdasarkan pada jenis kelamin mengingat berdasarkan hasil penelitian menunjukkan bahwa pada usia sekolah dasar, siswa laki-laki dan perempuan dapat mencapai prestasi yang sama, tetapi pada usia sekolah yang lebih

tinggi, prestasi belajar yang dicapai siswa laki-laki lebih baik dari pada yang dicapai siswa perempuan (Phelps-Borrowman, 2002: 3, Jacobs, 2002: 2, Tobias, 2002: 2).

Permasalahan dalam penelitian ini dirumuskan sebagai berikut: (1) Bagaimana karakteristik perangkat soal Matematika Ujian Nasional SLTP di Propinsi Daerah Istimewa Yogyakarta tahun pelajaran 2004/2005 ditinjau dari bias tes (keberfungsian tes diferensial) berdasarkan teori respons-butir?, dan (2) Bagaimana karakteristik butir perangkat soal Matematika Ujian Nasional SLTP di Propinsi Daerah Istimewa Yogyakarta tahun pelajaran 2004/2005 ditinjau dari bias butir (keberfungsian butir diferensial) berdasarkan teori respons-butir? Penelitian ini cukup penting karena masih jarang dilakukan penelitian tentang kegiatan pendeteksian bias tes dan bias butir pada perangkat soal yang diujikan baik dalam skala lokal maupun nasional. Oleh karena itu, secara teoretis dan akademis, hasil penelitian ini diharapkan dapat digunakan sebagai dasar atau pendukung kesimpulan awal bagi para peneliti lain untuk melakukan penelitian lanjutan (pengembangan) maupun penelitian sejenis (perluasan). Adapun secara praktis penelitian ini diharapkan dapat bermanfaat bagi segenap pengambil kebijakan di bidang pengukuran dan pendidikan serta para peneliti untuk bidang penelitian.

B. Kajian Pustaka

1. Evaluasi dalam Pendidikan

Gronlund (1985: 5) yang menyatakan bahwa evaluasi dalam pendidikan merupakan suatu proses sistematis guna mengumpulkan, menganalisis, dan mendapatkan informasi yang jelas untuk menentukan efektivitas kegiatan pembelajaran yang dilakukan pada saat program berlangsung maupun di akhir pelaksanaan program. Berkaitan dengan pembelajaran Matematika, evaluasi merupakan suatu alat untuk mengimplementasikan berbagai standar dan mengefektifkan perubahan secara sistematis. Fokus kegiatan evaluasi adalah pada apa yang terjadi di dalam kelas yang berupa interaksi para guru dan siswa (Webb, 1989: 189). Hal ini berarti, kegiatan evaluasi dilakukan secara menyeluruh terhadap berbagai aspek yang di antaranya meliputi aspek kognitif, afektif, dan psikomotorik.

Mengenai evaluasi terhadap aspek kognitif dalam pendidikan khususnya pembelajaran Matematika, akan memerlukan instrumen yang diantaranya berupa tes. Tes dapat diklasifikasikan dalam beberapa macam tergantung dari bentuk, tipe, dan ragamnya (Zainul & Nasution, 2001: 4). Adapun fungsi tes hasil belajar sebagaimana dikemukakan Gronlund (1985: 16) adalah sebagai alat untuk penempatan, fungsi formatif, fungsi diagnostik, dan fungsi sumatif. Berdasarkan bentuknya tes hasil belajar dapat dikelompokkan menjadi tiga jenis yaitu: (1) objektif, yang sederhana terdiri dari bentuk jawaban singkat, benar-salah atau dua pilihan, dan menjodohkan, (2) objektif pilihan ganda dengan alternatif jawaban lebih dari dua, dan (3) uraian (Gronlund, 1985: 144). Pada tes hasil belajar Matematika yang berupa Ujian Nasional untuk siswa SLTP, tes yang digunakan berbentuk objektif pilihan ganda dengan empat alternatif pilihan jawaban dan disiapkan oleh Pusat Penilaian Pendidikan.

2. Teori Tes

a. Teori Tes Klasik

Teori tes klasik atau teori skor-murni klasik (Allen & Yen, 1979: 57-60) didasarkan pada suatu model aditif yakni skor amatan yang merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Skor kesalahan pengukuran diasumsikan tidak berkorelasi dengan skor sebenarnya dan skor kesalahan pada tes-tes yang lain untuk peserta tes yang sama, dan rata-rata dari skor kesalahan ini sama dengan nol. Asumsi-asumsi tersebut menjadi dasar dalam mengembangkan formula-formula untuk menentukan validitas dan reliabilitas tes serta indeks kesukaran dan daya pembeda.

Allen & Yen (1979: 121) menyatakan bahwa secara umum indeks kesukaran suatu butir sebaiknya terletak pada interval 0.3 – 0.7 karena pada interval ini informasi tentang kemampuan siswa akan diperoleh secara maksimal. Namun demikian, besarnya indeks kesukaran suatu perangkat soal dapat dirancang dengan mempertimbangkan tujuan penyusunan perangkat soal tersebut. Menurut Allen & Yen (1979: 122) menyatakan daya pembeda suatu butir soal (d) merupakan suatu parameter yang memberikan informasi tentang seberapa besar daya soal itu untuk dapat membedakan peserta tes yang berkemampuan tinggi dengan peserta tes yang berkemampuan rendah. Penentuan indeks daya pembeda suatu butir soal dapat dilakukan dengan tiga cara yaitu: (1) indeks

diskriminasi, (2) indeks korelasi *biserial/point biserial*, dan (3) indeks keselarasan. Suatu butir soal dikatakan baik jika daya pembedanya lebih besar atau sama dengan 0.2.

b. Teori Respons-Butir

Oleh karena teori tes klasik memiliki berbagai keterbatasan dengan adanya sifat *group dependent* dan *item dependent* maka munculnya teori respons-butir menjadi sangat berguna dan terus dikembangkan karena mampu mengatasi keterbatasan tersebut (Hambleton, Swaminathan & Rogers, 1991: 2-5). Teori respons-butir dikembangkan atas dasar dua postulat yaitu: (1) performansi subyek pada suatu item dapat diprediksikan oleh seperangkat faktor yang disebut *latent trait* atau kemampuan, dan (2) hubungan pada performansi subyek pada suatu item dan perangkat kemampuan laten yang mendasarinya digambarkan oleh fungsi naik monoton yaitu *Item Characteristic Curve (ICC)*.

Secara umum ciri-ciri teori-respons butir adalah sebagai berikut: (1) karakteristik butir tidak tergantung pada peserta ujian, (2) skor yang digambarkan peserta ujian tidak tergantung pada tes, (3) merupakan model yang lebih menekankan pada tingkat butir daripada tingkat tes, (4) merupakan model yang tidak mensyaratkan secara ketat tes paralel untuk menaksir reliabilitas, dan (5) merupakan model yang menguraikan sebuah ukuran keputusan untuk tiap skor kemampuan yakni ada hubungan fungsional antara peserta ujian terhadap tingkat kemampuan yang dimiliki. Adapun asumsi-asumsi yang melandasi teori-respons butir adalah unidimensi, independensi lokal, dan fungsi karakteristik butir atau kurva karakteristik butir. Unidimensi artinya bahwa dimensi karakter peserta yang diukur oleh suatu tes itu tunggal. Independensi lokal adalah bahwa apabila kemampuan-kemampuan yang mempengaruhi kinerja tes dianggap konstan maka respons subjek terhadap setiap butir soal secara statistik tidak saling terkait. Fungsi karakteristik butir atau kurva karakteristik butir merupakan suatu pernyataan matematik yang berkaitan dengan probabilitas sukses memberikan respons yang benar pada suatu butir dengan kemampuan yang diukur dengan karakteristik butir (Hambleton, Swaminathan & Rogers, 1991: 13).

Teori respons-butir juga mempunyai suatu sifat yang disebut invariansi parameter butir dan parameter kemampuan (*invariance of item and ability parameter*). Sifat ini memberikan pengertian bahwa parameter yang mengkarakteristikkan suatu butir tes tidak

tergantung pada distribusi kemampuan peserta dan parameter yang mengkarakteristikan kemampuan peserta tidak tergantung pula pada karakter tes yang digunakan.

Pada umumnya dalam teori respons-butir digunakan model distribusi logistik. Hal ini dikarenakan model distribusi logistik lebih sederhana analisis matematikanya (Mardapi, 1991: 4). Ada tiga macam model logistik dalam teori respon butir yaitu model logistik satu parameter, model logistik dua parameter, dan model logistik tiga parameter. Model logistik satu parameter ditentukan oleh satu karakteristik butir yaitu tingkat kesukaran. Model logistik dua parameter ditentukan oleh dua karakteristik butir yaitu tingkat kesukaran dan daya pembeda. Model logistik tiga parameter ditentukan oleh tiga karakteristik butir yakni tingkat kesukaran, daya pembeda, dan *pseudo guessing* (faktor tebakan). Ketiganya berlaku untuk butir dengan respons yang bersifat dikotomi yaitu butir yang skornya benar dan salah. Pada model logistik tiga parameter dengan adanya faktor tebakan memungkinkan subjek dengan kemampuan yang rendah mempunyai peluang untuk menjawab butir soal dengan benar. Pada tes kemampuan bentuk pilihan ganda berpeluang besar bagi peserta untuk melakukan tebakan jawaban dalam merespons sehingga penggunaan model logistik tiga parameter merupakan model yang paling tepat.

Penerapan teori respons-butir model logistik dalam analisis butir soal suatu tes selain menggunakan tiga parameter butir yaitu tingkat kesukaran (b), daya beda (a), dan faktor tebakan (c), terdapat pula satu parameter peserta yaitu parameter kemampuan (θ). Penaksiran parameter menggunakan rumus perhitungan matematik yang cukup rumit, apalagi bila melibatkan sebuah tes dengan sejumlah butir yang cukup banyak yang direspons sejumlah peserta yang cukup banyak juga. Model logistik tiga parameter menaksir tiga parameter butir sehingga memerlukan waktu yang cukup lama untuk menyelesaikan seluruh perhitungan tersebut. Untuk mengatasi kendala tersebut sejak tahun 1979 telah dikembangkan beberapa program komputer seperti: *Bical*, *Logist*, *Bilog*, *Nohram*, *Microscale*, *Rascal*, *Ascal*, dan *Rida* (Hambleton, et. Al., 1991: 7).

Fungsi informasi tes dalam teori respons-butir merupakan penjumlahan dari fungsi informasi seluruh butir pada tes tersebut pada tingkat kemampuan θ (Hambleton, Swaminathan & Rogers, 1985: 94). Fungsi informasi dengan kesalahan penaksiran standar (*SEM*) mempunyai hubungan yang berbanding terbalik kuadratik (Hambleton, Swaminathan, Rogers, 1991: 94).

3. Teori Bias Tes (*Differential Test Functioning/DTF*) dan Bias Butir (*Differential Item Functioning/DIF*)

a. Bias Tes dan Bias Butir

Bias dari suatu tes merupakan suatu kondisi tes yang tidak adil (*unfair*), tidak konsisten, dan terkontaminasi oleh faktor-faktor di luar faktor yang hendak dites (Osterlind, 1983: 10). Sebuah butir soal dikatakan tidak bias jika probabilitas keberhasilan pada butir soal tersebut adalah sama untuk pengambil tes yang memiliki kemampuan sama dari populasi yang sama tanpa memperhatikan keanggotaan kelompok mereka (Camilli & Shepard, 1994: 62). Holland & Thayer (Camilli & Shepard, 1994: 16) menyebut bias butir sebagai keberfungsian butir diferensial (*Differential Item Functioning/DIF*). Adapun bias tes disebut sebagai keberfungsian tes diferensial (*Differential Test Functioning/DTF*). Baik *DIF* maupun *DTF* dapat dideteksi lewat berbagai prosedur baik melalui pendekatan teori tes klasik maupun teori respons-butir.

b. Metode Pendeteksian *DTF* dan *DIF* Berdasarkan Teori Respons-Butir

Pendeteksian bias tes (*DTF*) dapat dilakukan berdasarkan teori respons-butir yaitu dengan melihat grafik nilai peluangnya. Adapun untuk pendeteksian bias butir (*DIF*) pada suatu tes berdasarkan teori respons butir dapat dilakukan dengan berbagai pendekatan yaitu: (1) uji perbedaan parameter kesukaran butir, (2) metode penyimpangan butir, (3) uji kai kuadrat Lord, (4) distribusi sampling empirik indeks *DIF*, dan (5) perbandingan model teori respons-butir. Pendekatan perbandingan model teori respons-butir dibagi menjadi empat kategori utama yaitu: (a) umum, (b) log linier, (c) informasi terbatas, dan (d) teori respons-butir 3 parameter (Camilli & Shepard, 1994: 70-80).

Pada penelitian ini difokuskan pada penggunaan grafik nilai peluang untuk mendeteksi *DTF* dan metode perbandingan model teori respons-butir 3 parameter untuk mendeteksi *DIF* berdasarkan perbedaan gender. Langkah-langkah pendeteksian *DIF* dengan metode ini diuraikan dalam tujuh langkah sebagai berikut (Camilli & Shepard, 1994: 79): (1) Menaksir parameter butir dengan teori respons-butir model logistik tiga parameter sehingga diperoleh statistik kecocokan yang mendekati distribusi kai kuadrat $G(1)$ untuk tes yang terdiri dari K butir, (2) Menentukan salah satu butir tes misalnya butir i untuk dievaluasi, (3) Butir tes dibuat seolah-olah menjadi dua butir dengan butir tes pertama berisi jawaban dari salah satu kelompok, misalnya A dan tidak direspons oleh

kelompok lainnya, misalnya B dan butir kedua berisi jawaban dari kelompok B yang tidak direspons oleh kelompok A, (4) Menaksir kembali parameter dan didapatkan harga kai kuadrat transformasi rasio kebolehhadian $G(2)$ untuk tes yang terdiri dari $K+1$ butir, (5) Menghitung selisih nilai distribusi kai kuadrat dengan persamaan sebagai berikut:

$$\chi^2_{hitung} = G(1) - G(2) \dots\dots\dots(1)$$

Harga yang diperoleh tersebut selanjutnya dibandingkan dengan harga kritik distribusi kai kuadrat dengan derajat kebebasan 3 dan taraf signifikansi α , (6) Jika harga χ^2_{hitung} lebih besar daripada harga χ^2_{tabel} berarti pada butir i menunjukkan adanya *DIF* yang secara statistik signifikan pada taraf signifikansi α , dan (7) Mengulangi langkah kedua sampai dengan keenam untuk mengevaluasi *DIF* pada butir-butir yang lain.

Perbedaan parameter menyebabkan *DIF* terjadi dalam dua kategori umum yaitu: (1) *DIF* konsisten atau *uniform* yang terjadi jika kurva karakteristik butir berbeda dan tidak saling berpotongan/bersilangan, dan (2) *DIF* tidak konsisten atau *non uniform* yang terjadi jika kurva karakteristik butir berbeda tapi berpotongan pada suatu titik pada skala θ (Camilli & Shepard, 1994: 64). Untuk mengetahui kelompok mana yang lebih diuntungkan, dapat dilihat secara langsung dari grafik nilai peluangnya.

4. Hakikat Matematika Sekolah

Matematika yang diajarkan di pendidikan dasar (SD dan SLTP) dan pendidikan menengah (SLTA) biasa disebut dengan Matematika Sekolah (Soedjadi, 2000: 37). Fungsi mata pelajaran Matematika adalah sebagai alat, pola pikir, dan ilmu atau pengetahuan (Suherman, 2003: 76). Oleh karena itu, Matematika Sekolah berperan penting bagi siswa untuk memberikan bekal pengetahuan dan pembentukan sikap serta pola pikir sebagai warga negara pada umumnya untuk dapat hidup layak baik untuk kemajuan negaranya maupun Matematika itu sendiri dalam rangka melestarikan dan mengembangkannya. Selain itu, sesuai dengan UUD 1945 pasal 31, pendidikan dan pengajaran merupakan hak setiap warga negara. Ini berarti bahwa setiap warga negara tanpa melihat suku, jenis kelamin, agama, ras, dan golongan berhak mendapatkan pendidikan dan pengajaran. Hal ini mengimplikasikan bahwa setiap siswa juga memiliki hak yang sama dalam mencapai prestasi belajar Matematika di sekolah.

5. Bank Soal

Penyusunan bank soal merupakan salah satu langkah untuk menunjang keberhasilan program penilaian dalam pendidikan. Bank soal disusun dan dikelola secara sistematis melalui langkah-langkah pengembangan yang standar atau baku. Langkah-langkah tersebut antara lain: (1) penyusunan kisi-kisi, (2) penulisan soal, (3) telaah (*review*) dan perbaikan (*revisi*), (4) perakitan tes, (5) ujicoba tes, (6) entri data, (7) analisis butir soal dan kalibrasi soal, (8) seleksi soal, dan (9) pengkartuan/komputerisasi bank soal (Tim Sisjian, 1997: 2).

Informasi yang terdapat dalam butir-butir bank soal akan sangat membantu dalam pengembangan tes. Informasi yang ada membantu dalam menyeleksi dan menetapkan butir soal mana yang akan dipilih jika akan menyusun tes dengan karakteristik butir tertentu. Hal ini akan meningkatkan kualitas tes dan hasil pengukuran karena tes yang disusun sesuai dengan tingkat kemampuan peserta sehingga informasi hasil tes menjadi tinggi. Informasi hasil pengukuran yang tinggi akan membantu dalam pengambilan kebijakan yang tepat. Berdasarkan kenyataan tersebut menunjukkan bahwa pengadaan bank soal baik di tingkat daerah maupun pusat sangat diperlukan di bidang pendidikan.

C. METODE PENELITIAN

1. Tempat dan Waktu Penelitian

Penelitian dilakukan di Kantor Dinas Pendidikan Propinsi DIY. Kegiatan penelitian dilaksanakan selama empat bulan yakni dari bulan Mei 2006 sampai dengan bulan Agustus 2006.

2. Jenis Penelitian

Penelitian ini bersifat *exploratif*. Penelitian ini akan mengungkap bias tes dan bias butir berdasarkan teori respons-butir.

3. Sumber Data

Data dalam penelitian ini adalah respons siswa peserta Ujian Nasional SLTP pada paket soal mata pelajaran Matematika kode P1 di Propinsi DIY tahun pelajaran 2004/2005. Sumber data berupa lembar jawaban siswa yang telah didokumentasi dengan menggunakan komputer.

4. Populasi dan sampel

Populasi penelitian ini adalah respons siswa peserta Ujian Nasional SLTP pada mata pelajaran Matematika di propinsi DIY tahun pelajaran 2004/2005 yang tersebar dalam 436 SLTP dengan keseluruhan peserta berjumlah 42.574 siswa. Sampel penelitian diambil dengan teknik *purposive random sistematik sampling* sebanyak 2000 respons siswa yang terdiri atas 1000 respons siswa laki-laki dan 1000 respons siswa perempuan.

5. Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi yaitu dengan mengutip respons siswa peserta Ujian Nasional SLTP pada mata pelajaran Matematika di Propinsi DIY tahun pelajaran 2004/2005. Data tersebut diperoleh di kantor Dinas Pendidikan Propinsi DIY.

6. Teknik Analisis Data, Cara Penafsiran, dan Penyimpulan Hasil Penelitian

Kegiatan analisis data dalam penelitian ini meliputi dua hal pokok yaitu: (1) pendeteksian *DTF* berdasarkan teori respons-butir dan (2) pendeteksian *DIF* dengan metode perbandingan model teori respons-butir 3 parameter. Pendeteksian *DTF* dilakukan melihat grafik nilai peluangnya dan pendeteksian *DIF* dilakukan dengan menggunakan metode perbandingan model teori respons-butir 3 parameter. Pendeteksian *DIF* dengan metode ini dilakukan dengan terlebih dahulu mengestimasi kemungkinan maksimum untuk keseluruhan butir tes atau $G(1)$ dilanjutkan estimasi kemungkinan maksimum untuk tiap butir atau $G(2)$. Harga χ^2_{hitung} dihitung berdasarkan persamaan 1 kemudian dibandingkan dengan χ^2_{tabel} dengan derajat kebebasan 3 dan taraf signifikansi 0,05. Untuk mengetahui kelompok mana yang lebih diuntungkan apabila ada butir yang terdeteksi mengandung *DIF* dilakukan dengan melihat grafik nilai peluangnya.

D. HASIL PENELITIAN DAN PEMBAHASAN

1. Deskripsi Penyelenggaraan UAN SLTP di Propinsi DIY

Pada Ujian Nasional tahun Pelajaran 2004/2005, untuk mata pelajaran Matematika dipersiapkan paket soal utama untuk siswa awas, paket soal utama untuk siswa tunanetra, dan paket soal susulan. Masing-masing paket soal terdiri atas satu bendel soal, satu bendel lembar jawaban, dan pedoman penilaian. Berdasarkan penjelasan pihak penyelenggara Ujian Nasional SLTP di propinsi DIY bahwa mengingat wilayah propinsi DIY yang relatif tidak luas maka perangkat soal yang digunakan sama untuk semua

peserta tes dan khusus untuk mata pelajaran Matematika SLTP digunakan paket soal dengan kode P1.

2. Deskripsi Hasil Penelitian

a. Karakteristik Perangkat Soal UAN SLTP Mata Pelajaran Matematika di Propinsi DIY Tahun Pelajaran 2004/2005

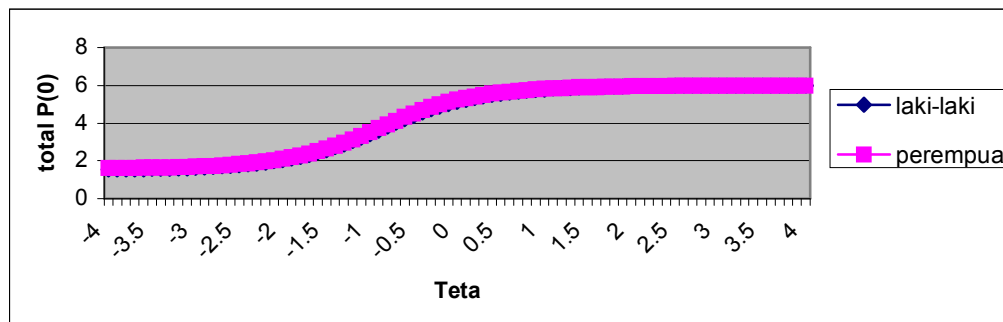
Pada penelitian ini, karakteristik perangkat soal Matematika Ujian Nasional SLTP berdasarkan teori tes klasik dilakukan dengan bantuan program *BiLog* versi 3.07 fase pertama. Berdasarkan hasil analisis menunjukkan bahwa dari 30 butir soal yang dianalisis seluruhnya termasuk kategori baik. Adapun karakteristik perangkat soal berdasarkan teori respons-butir dilakukan dengan menggunakan program *BiLog* versi 3.07 model logistik tiga parameter fase kedua. Sedangkan untuk fungsi informasi butir dihitung dengan menggunakan program EXCEL 2000. Berdasarkan hasil analisis menunjukkan bahwa fungsi informasi tes Matematika Ujian Nasional ini adalah sebesar 9.469356 dan nilai ini tercapai jika kemampuan siswa sebesar 0.35. Hasil analisis juga menunjukkan bahwa dari 30 butir soal yang dianalisis terdapat 9 butir tes yang tidak cocok dengan model dan 21 butir tes yang cocok dengan model. Butir tes yang cocok dengan model terdiri atas 20 butir tes dengan kualitas baik dan 1 butir tes dengan kualitas cukup baik. Adapun besarnya kesalahan pengukuran dalam penelitian ini dihitung berdasarkan teori respons butir yaitu sebesar 0.324967.

2. Pendeteksian *DTF* dan *DIF*

Pendeteksian *DTF* dalam penelitian ini dilakukan dengan melihat grafik nilai peluangnya. Dari hasil analisis diperoleh grafik nilai peluang untuk keseluruhan butir tes adalah sebagai berikut:

Gambar 1.

Grafik Nilai Peluang Seluruh Butir Tes



Berdasarkan grafik fungsi peluang di atas, tampak bahwa pada perangkat soal Matematika Ujian Nasional SLTP memuat *DTF uniform* yang sedikit menguntungkan siswa perempuan dibandingkan siswa laki-laki. Adapun untuk hasil pendeteksian *DIF* menggunakan metode perbandingan model teori respons-butir 3 parameter menunjukkan bahwa pada taraf signifikansi 0,01 terdapat 6 butir soal yang terdeteksi memuat *DIF* secara statistik yaitu butir soal nomor 2, 8, 11, 15, 18, dan 25. Apabila dilihat grafik nilai peluang untuk tiap butir soal yang terdeteksi memuat *DIF* menunjukkan bahwa: butir nomor 2 memuat *DIF non uniform* yang menguntungkan siswa perempuan pada skala kemampuan kurang dari -2.6 dan menguntungkan siswa laki-laki pada skala kemampuan lebih dari -2.6 , butir nomor 8 memuat *DIF non uniform* yang menguntungkan siswa perempuan pada skala kemampuan lebih dari $1,8$ dan menguntungkan siswa laki-laki pada skala kemampuan kurang dari $-1,8$, butir nomor 11 memuat *DIF non uniform* yang menguntungkan siswa perempuan pada skala kemampuan lebih dari -2.0 dan menguntungkan siswa laki-laki pada skala kemampuan lebih dari -2.0 , butir nomor 15 memuat *DIF non uniform* yang menguntungkan siswa perempuan pada skala kemampuan kurang dari $-1,8$ dan menguntungkan siswa laki-laki pada skala kemampuan lebih dari $-1,8$, butir nomor 18 memuat *DIF non uniform* yang menguntungkan perempuan pada skala kemampuan kurang dari 0 dan menguntungkan laki-laki pada skala kemampuan lebih dari 0 , dan butir nomor 25 memuat *DIF non uniform* yang menguntungkan perempuan pada skala kemampuan kurang dari 0 dan menguntungkan laki-laki pada skala kemampuan lebih dari 0 .

3. Pembahasan

Berdasarkan hasil analisis menurut teori tes klasik yang dilakukan dengan bantuan program *Bilog* versi 3.07 fase pertama menunjukkan bahwa dari 30 butir soal yang dianalisis seluruhnya termasuk kategori baik. Dengan demikian, berdasarkan teori tes klasik ini, seluruh butir soal dapat dipilih dalam kegiatan pengembangan bank soal. Adapun berdasarkan teori respons-butir, pada taraf signifikansi 0.05 terdapat 9 butir soal yang tidak cocok dengan model sehingga termasuk kategori butir soal yang tidak baik yaitu butir soal nomor 1, 3, 6, 7, 10, 13, 14, 16, dan 29. Hasil analisis juga menunjukkan bahwa dari 30 butir soal yang dianalisis terdapat 21 butir tes yang cocok dengan model. Butir tes yang cocok dengan model terdiri atas 20 butir dengan kualitas baik (nomor

2,5,8,9,11,12,15,16,17,18,19,20,21,22,23,24,26,27,28, dan 30) serta 1 butir soal dengan kualitas cukup baik (nomor 4). Dengan demikian, berdasarkan teori respons butir ini, apabila akan dipilih butir soal untuk pengembangan bank soal ada 21 butir soal yang dapat dipilih.

Berdasarkan teori respons-butir, perangkat soal Ujian Nasional SLTP mata pelajaran Matematika ini memiliki fungsi informasi tes sebesar 9.469356 dan nilai ini tercapai jika kemampuan siswa sebesar 0.35. Hal ini mengindikasikan bahwa tes ini dapat mengungkap kemampuan siswa secara maksimum pada peserta tes yang kemampuannya 0.35. Selanjutnya mengenai kesalahan pengukuran, berdasarkan hasil analisis menunjukkan bahwa besarnya kesalahan pengukuran dengan teori respons butir sebesar 0.324967. Kesalahan pengukuran digunakan untuk memahami kesalahan yang bersifat acak yang memengaruhi skor peserta dalam pelaksanaan tes. Pada perangkat soal Ujian Nasional SLTP mata pelajaran Matematika ini, misalkan siswa ke- i memperoleh skor X_i . Berdasarkan teori respons-butir diperoleh bahwa kesalahan pengukurannya sebesar 0.324967. Apabila skor total maksimum yang dapat dicapai siswa adalah 30 maka skor sebenarnya yang diperoleh siswa berkisar pada $X_i \pm 0.324967$.

Pendeteksian *DTF* dalam penelitian ini dilakukan dengan melihat grafik nilai peluangnya. Dari hasil penelitian ini diperoleh bahwa perangkat soal Matematika Ujian Nasional di Propinsi DIY tahun pelajaran 2004/2005 memuat *DTF uniform* yang sedikit menguntungkan siswa perempuan. Adapun pendeteksian *DIF* dengan metode perbandingan model teori respons-butir menemukan adanya 6 butir soal yang terdeteksi memuat *DIF* berdasarkan perbedaan gender yaitu butir soal nomor 2, 8, 11, 15, 18, dan 25. Terdeteksinya *DIF* pada perangkat soal ini menunjukkan bahwa perbedaan gender memang dapat menjadikan suatu butir soal menjadi bias. Sebagai contoh, butir soal nomor 2 terindikasi memuat *DIF*. Hal ini berarti bahwa pada butir soal nomor 2 ini dua siswa dengan kemampuan sama yang berada dalam kelompok yang berbeda memberikan respons berbeda sehingga mendapatkan hasil berbeda pula. Hal ini tentu saja menjadikan ketidakadilan terhadap salah satu kelompok siswa dan menunjukkan adanya sifat tidak fair dalam pengukuran yang semestinya dihindari oleh penyusun perangkat soal serta perlu dikaji lebih lanjut. Berdasarkan grafik nilai peluang menjawab benar tiap kelompok peserta tes dengan parameter butir yang diperoleh dari hasil estimasi untuk tiap kelompok

peserta tes menunjukkan bahwa butir soal nomor 2 ini memuat *DIF non uniform* yang menguntungkan siswa perempuan pada skala kemampuan kurang dari -2.6 dan menguntungkan siswa laki-laki pada skala kemampuan lebih dari -2.6 .

Terindikasinya *DIF* pada butir-butir soal tersebut menunjukkan adanya respons berbeda yang memberikan hasil berbeda pada dua siswa dengan kemampuan sama yang berada dalam kelompok berbeda yakni kelompok laki-laki dan kelompok perempuan. Adanya *DIF* pada suatu butir soal dapat menjadi petunjuk adanya suatu bias butir dan adanya *DTF* pada suatu tes dapat menjadi petunjuk adanya bias pada suatu tes. Namun untuk menentukan penyebab baik bias tes maupun bias butir masih membutuhkan penelitian lebih lanjut dan mendalam. Hal ini disebabkan *DTF* dan *DIF* pada suatu perangkat soal dapat disebabkan oleh banyak faktor seperti kemampuan guru, sarana dan prasarana, dan metode pengajaran. Penyelenggaraan tes juga dapat menjadi penyebab timbulnya *DTF* dan *DIF* terutama berkaitan dengan pengawasan tes dan tata tertib pelaksanaan suatu tes. Proses pembelajaran juga dapat memberikan pengaruh berbeda pada dua kelompok kemampuan yang sama. Sebagai contoh butir soal nomor 2 yang menurut metode perbandingan model teori respons-butir terindikasi memuat *DIF* berdasarkan perbedaan gender. Butir soal nomor 2 meminta siswa untuk menghitung harga pembelian. Bagi siswa yang memahami konsep untung dan rugi, hafal rumusnya, trampil dalam perhitungan, mampu memahami maksud soal kemungkinan besar akan mampu menyelesaikan soal dengan benar. Sedangkan bagi siswa yang tidak sering latihan dalam proses pembelajarannya walaupun mempunyai kemampuan yang sama dapat memberikan respons yang salah. Peneliti tidak dapat memastikan apakah terindikasinya *DTF* dan *DIF* tersebut disebabkan oleh proses pembelajaran atau yang lainnya karena untuk menyatakan penyebabnya diperlukan kajian yang lebih mendalam.

Adanya *DTF* dan *DIF* pada suatu tes dapat pula disebabkan oleh pengawasan dan tata tertib pelaksanaan tes yang kadang diberlakukan berbeda untuk tiap rayon. Pengawasan yang kadang tidak ketat akan menyebabkan peserta ujian lebih leluasa untuk bekerjasama dengan sesama peserta ujian yang lain. Karena ada kemungkinan peserta bekerjasama dengan peserta lain maka akan menghasilkan estimasi parameter yang berbeda dengan harga parameter butir yang sebenarnya. Demikian pula tata tertib pelaksanaan tes yang diberlakukan tidak sama pada masing-masing rayon penyelenggara,

misalnya penggunaan kalkulator sebagai alat bantu hitung walaupun dalam naskah butir tes tidak boleh menggunakan kalkulator. Keadaan demikian tentunya merugikan peserta ujian rayon lain yang menerapkan tata tertib lebih ketat sehingga dua orang siswa dengan kemampuan sama bisa memberikan respons berbeda sehingga memperoleh hasil yang berbeda pula. Perbedaan ini tentu saja merugikan salah satu kelompok yang perlu untuk dihindari. Namun demikian, masih banyak faktor-faktor lain yang dapat menyebabkan terjadinya DIF pada suatu butir soal yang memang harus dikaji lebih lanjut dan mendalam.

5. SIMPULAN DAN SARAN

a. Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat diambil simpulan sebagai berikut: (1) Karakteristik perangkat soal Matematika Ujian Nasional SLTP di Propinsi DIY tahun pelajaran 2004/2005 adalah sebagai berikut: (a) Berdasarkan hasil analisis menggunakan teori tes klasik, secara keseluruhan 30 butir soal dapat direkomendasikan untuk pengembangan bank soal berdasarkan teori tes klasik, (b) Berdasarkan hasil analisis menggunakan teori respons-butir menunjukkan bahwa pada perangkat soal terdapat 9 butir soal dengan kualitas tidak baik, 20 butir soal dengan kualitas baik, dan 1 butir soal dengan kualitas cukup baik, (c) besarnya fungsi informasi tes adalah 9,469356 yang dicapai pada kemampuan siswa 0.35, dan (d) kesalahan pengukuran perangkat soal Matematika Ujian Nasional SLTP berdasarkan teori respons-butir sebesar 0,324967. (2) Perangkat soal Matematika Ujian Nasional SLTP di Propinsi DIY tahun pelajaran 2004/2005 ini memuat bias tes (*DTF*) *uniform* yang sedikit menguntungkan siswa perempuan. (3) Berdasarkan hasil analisis pendeteksian *DIF* menggunakan metode perbandingan model teori respons-butir menunjukkan bahwa pada perangkat soal Matematika Ujian Nasional SLTP di Propinsi DIY tahun pelajaran 2004/2005 terdapat 6

butir soal yang terdeteksi memuat *DIF* berdasarkan perbedaan gender yaitu butir soal nomor 2, 8, 11, 15, 18, dan 25.

b. Saran

Berbagai saran yang diajukan berdasarkan hasil dari penelitian ini antara lain adalah: (1) Kepada penyusun perangkat soal Matematika SLTP baik dalam skala lokal maupun nasional sebaiknya menggunakan soal-soal yang secara empiris terbukti berkualitas baik dan tidak memuat *DTF* maupun *DIF*, (2) Pembuktian secara empiris terhadap kualitas suatu butir soal perlu lebih dikembangkan dan disosialisasikan oleh berbagai praktisi pendidikan misalnya dengan mengadakan berbagai pelatihan bagi guru-guru baik bidang studi Matematika maupun lainnya untuk dapat melakukan analisis empiris terhadap perangkat tes yang digunakan, (3) Untuk mendeteksi *DIF* dan *DTF* suatu perangkat soal lainnya yang bersifat setara dengan Ujian Nasional dapat dilakukan berdasarkan teori respons-butir, (4) Untuk lebih memperluas kajian tentang *DTF* dan *DIF* perlu adanya penelitian lebih lanjut dengan menggunakan data dan metode yang berbeda.

DAFTAR PUSTAKA

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Belmont, CA: Wadsworth, Inc.
- Bolt, D. M. (2000). A SIBTEST approach to testing DIF hypotheses using experimentally designed test items. *Journal of Educational Measurement*, 37(4), 307-327.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying bias test items*. Thousand Oaks, CA: Sage Publication.
- Crane, P., Belle, G., & Larson, E. (2004). *Differential item functioning in a cognitive test*. Diambil pada tanggal 27 April 2004, dari <https://www.alz.washington.edu/NONMEMBER/KENTUCKY0502/CRANE/TSLD011>.
- Crocker, L. & Algina J. (1986). *Introduction to classical and modern test theory*. New York: CBS college publishing.
- Gierl, M., Khaliq, S. N., & Boughton, K. (Juli 1999). Gender differential item functioning in mathematics and science: prevalence and policy implications. *Makalah dalam simposium Improving Large Scale*

- Assesment in Education pada pertemuan tahunan the Canadian Society for the Study of Education*, Quebec, Canada. Diambil pada tanggal 14 Juli 2002, dari <http://www.ncrel.org/sdrs/>.
- Gronlund, N. E. (1976). *Measurement and evaluating in teaching*. New York: Macmillan Publishing Co.
- Hambleton, R. K., Swaminathan H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Hamilton, L. S. & Snow, R. E. (1998). *Exploring differential item functioning on science achievement test*. Los Angeles: Center for the study of Evaluation National Centre for Research of Evaluation. Diambil pada tanggal 20 Agustus 2002, dari <http://www.cse.ucla.edu/cresst/reports/TECH483.pdf>.
- Hullin, C. L., et al. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones–Irwin.
- Mardapi, D. (1991). Konsep dasar teori respons butir: Perkembangan dalam bidang pengukuran pendidikan. *Cakrawala Pendidikan* 3(X). 1-16.
- Mislevy, R. J., & Bock, R. D. (1990). *BLOG 3: Item analysis & test scoring with binary logistic models* (computer program). Mooresville, IN: Scientific Software, Inc.
- Osterlind, S. J. (1983). *Test item bias*. Beverly Hills, CA: Sage Publication Inc.
- Scrams, D.J. & Mcleod, L.D. (2000). An expected response function approach to graphical differential item functioning. *Journal of Educational Measurement*, 37(3), 263-280.
- Soedjadi, R. (2000). *Kiat pendidikan matematika di Indonesia: konstataasi keadaan masa kini menuju harapan masa depan*. Jakarta: Dirjen Dikti.
- Stark, S. et al. (2001). IRT Modeling Lab. *Detection of differential item/test functioning (DIF/DTF) using IRT*. University of Illinois at Urbana Champaign online. Diambil pada tanggal 20 Maret 2003, dari <http://www.work.psych.uiuc.edu/irt/>.
- Stark, S. & Chernyshenko, O. (2002). *Detection of differential item/test functioning (DIF/DTF) using IRT*. University of Illinois at Urbana Champaign online. Diambil pada tanggal 3 Januari 2003, dari <http://www.work.psych.uiuc.edu/irt/>.
- Sub Dinas PLB dan DIKDAS. (2002). *Laporan hasil ujian akhir nasional SLTP tahun pelajaran 2004/2005*. Yogyakarta: Dinas Pendidikan Propinsi DIY.
- Suherman, E. et al. (2003). *Strategi pembelajaran matematika kontemporer*. Bandung: FMIPA UPI.
- Tim Sisjian. (1997). *Bank soal*. Jakarta: Pusat Penelitian dan Pengembangan Sistem Pengujian.
- Wright, B. D. & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.
- Zainul, A. & Nasution, N. (2001). *Penilaian hasil belajar*. Jakarta: PAU-PPAI, Universitas Terbuka.