# KEBERFUNGSIAN ITEM DIFFERENSIAL PADA PERANGKAT TES UJIAN NASIONAL MATEMATIKA SEKOLAH MENENGAH ATAS DI JAWA TENGAH

Samsul Hadi
samsul2340@yahoo.co.id

## A. Pendahuluan

Ujian akhir secara nasional dilaksanakan sebagai tes hasil pembelajaran selama pendidikan pada tingkat Sekolah Menengah Atas (SMA) pada akhir tahun pelajaran 2002/2003, diatur pemerintah dengan Keputusan Menteri Pendidikan Nasional Nomor 153/U/2003. Keputusan ini menyebutkan bahwa untuk tahap akhir suatu jenjang pendidikan diselenggarakan suatu evaluasi tersendiri yang dinamakan UAN (Ujian Akhir Nasional) yang selanjutnya disebut Ujian Nasional. Pelaksanaannya dilakukan secara serempak untuk masing-masing tingkatan dengan pengaturan jadwal dan pengalokasian waktu yang seragam pula.

Tujuan UN sebagaimana tercantum dalam pasal 2 ayat 1 Surat Keputusan Mendiknas adalah 1) mengukur pencapaian hasil belajar peserta didik; 2) mengukur mutu pendidikan di tingkat nasional, provinsi, kabupaten, /kota, dan sekolah; 3) mempertanggungjawabkan penyelenggaraan pendidikan secara nasional, provinsi, kabupaten/kota, sekolah dan masyarakat.

Prosedur awal untuk mengetahui terdapat tidaknya bias item pada suatu item tes dilakukan dengan analisis *Differential item functioning* (keberfungsian butir diferensial) yang selanjutnya disingkat *DIF*. Analisis *DIF* digunakan untuk mengidentifikasi seluruh item yang memiliki perbedaan fungsi untuk kelompok yang berbeda. Suatu item tes menunjukkan *DIF* jika siswa-siswa yang mempunyai kemampuan yang sama, tetapi berasal dari kelompok yang berbeda, tidak dapat mempunyai peluang yang sama untuk menjawab benar. (Hambleton, et. *al*., 1991). Prosedur selanjutnya untuk menentukan terjadinya bias atau tidak pada suatu butir adalah dengan analisis logik seperti halnya mengapa item-item tes tampak lebih sulit untuk satu kelompok dibanding kelompok lain. Hanya jika suatu item relatif lebih sulit untuk satu kelompok dan kesulitan tersebut tidak relevan terhadap konstruk tes maka item tes tersebut dikatakan bias. Dengan demikian suatu item tes yang mengandung *DIF* tidak otomatis item tersebut bias karena masih banyak prosedur lain yang digunakan untuk menentukan bias atau tidaknya suatu item tes termasuk analisis logik dari para ahli bidang studi.

Konstruk UN matematika secara teoritik tentu didesain untuk mengukur satu dimensi yaitu dimensi kemampuan matematika. Akan tetapi, bisa saja dalam konstruk matematika tersebut mengandung dimensi ke dua. Dimensi kedua ini disebut sebagai pembantu atau *auxiliary* jika berhubungan dengan konstruk dan disebut pengganggu atau *nuisance* jika tidak berhubungan dengan konstruk (Roussos dan Stout dalam Gierl dkk., 2003). Item-item yang mengandung *DIF* karena terdapat dimensi pembantu tidak bersifat merugikan (Douglas dkk., 1996) .

Wilayah Propinsi Jawa Tengah yang terdiri dari 35 Dati II mempunyai cakupan wilayah yang cukup luas dengan beraneka macam perbedaan baik perbedaan kehidupan sosial, ekonomi, budaya, dan politik. Perbedaan tersebut menyebabkan latar belakang siswa yang mengikuti UN di tiap rayon penyelenggara UN tidak sama. Keadaan demikian sangat potensial terdapat *DIF* pada item-item tes UN yang disebabkan latar belakang siswa tiap-tiap rayon penyelenggara UN.

Perolehan rata-rata UN yang berbeda dari masing-masing rayon dapat menimbulkan masalah. Masalah itu adalah apakah perbedaan tersebut disebabkan perbedaan kemampuan atau karena adanya bias item yang menguntungkan bagi siswa yang

berada pada rayon-rayon tertentu. Terjadinya bias item ini sangat dimungkinkan karena latar belakang siswa pada rayon-rayon kotamadya yang merupakan daerah perkotaan akan berbeda dengan latar belakang siswa pada rayon-rayon kabupaten yang lebih banyak merupakan daerah pedesaan.

Untuk mengetahui apakah terjadi bias item pada item-item tes UN Matematika di Jawa Tengah tahun pelajaran 2011/2012 maka perlu diadakan penelitian *DIF* yang didasarkan pada perbedaan wilayah UN.

## B. Kajian Teori

### 1. Evaluasi

Pengukuran, penilaian dan evaluasi merupakan kegiatan yang bersifat hierarki. Artinya ketiga kegiatan tersebut dalam kaitannya dengan proses belajar mengajar tidak dapat dipisahkan satu sama lain dan dalam pelaksanaannya harus dilaksanakan secara berurutan. Evaluasi juga diartikan salah satu rangkaian kegiatan dalam meningkatkan kualitas, kinerja atau produktivitas suatu lembaga dalam melaksanakan tugasnya (Mardapi, 2008:8).

Sesuai dengan pengertian tersebut maka setiap kegiatan evaluasi atau penilaian merupakan suatu proses yang sengaja direncanakan untuk memperoleh informasi atau data, berdasarkan data tersebut kemudian dicoba membuat suatu keputusan. Sudah barang tentu informasi atau data yang dikumpulkan itu haruslah data yang sesuai dan mendukung tujuan evaluasi yang direncanakan.

### 2. Tujuan dan Fungsi Evaluasi dalam Pendidikan

Tujuan umum evaluasi pendidikan adalah untuk menghimpun bahan-bahan keterangan yang akan dijadikan sebagai bukti mengenai taraf perkembangan atau taraf kemajuan yang dialami oleh para peserta didik setelah mereka mengikuti proses pembelajaran dalam jangka waktu tertentu, mengetahui tingkat efektivitas dari metode-metode pembelajaran yang telah dipergunakan dalam proses pembelajaran selama jangka waktu tertentu.

Tujuan khusus evaluasi pendidikan adalah untuk merangsang kegiatan peserta didik dalam menempuh program pendidikan, untuk mencari dan menemukan faktor penyebab keberhasilan dan ketidakberhasilan peserta didik dalam mengikuti program pendidikan sehingga dapat dicari dan ditemukan jalan keluar atau cara-cara perbaikannya (Sudijono, 2006:17). Bagi pendidik, secara didaktik evaluasi pendidikan memiliki lima fungsi, yaitu:

a. Memberikan landasan untuk menilai hasil usaha (prestasi) yang telah dicapai oleh peserta didiknya,
b. Memberikan informasi yang sangat berguna untuk mengetahui posisi peserta didik dalam kelompoknya,
c. Memberikan bahan yang penting untuk memilih dan kemudian menetapkan status peserta didik,
d. Memberikan pedoman untuk mencari dan menemukan jalan keluar bagi peserta didik yang memang memerlukannya,
e. Memberikan petunjuk tentang sejauh manakah program pengajaran yang telah ditentukan dan telah dapat dicapai (Sudijono, 2006:12).

### 3. Pengertian Tes, Klasifikasi Tes, dan Tes Prestasi Belajar

Tes merupakan sejumlah pertanyaan yang memiliki jawaban yang benar atau salah. Tes diartikan juga sebagai sejumlah pertanyaan yang membutuhkan jawaban, atau sejumlah pertanyaan yang harus diberikan tanggapan dengan tujuan mengukur tingkat kemampuan seseorang atau mengungkap aspek tertentu dari orang yang dikenai tes

 ISBN 978-602-14215-0-5

(Mardapi, 2008:67). Cronbach dalam Azwar (2007:5) membagi tes menjadi dua kelompok besar, yaitu tes yang mengukur performansi maksimal *(maximum performance)* dan tes yang mengukur performansi tipikal *(typical performance).*

Tes yang mengukur performansi maksimal adalah jenis tes yang dirancang untuk mengungkap apa yang mampu dilakukan oleh seseorang dan seberapa baik ia melakukannya. Sedangkan tes yang mengukur performansi tipikal adalah jenis tes yang dirancang untuk mengungkap kecenderungan reaksi atau perilaku individu ketika berada dalam situasi situasi tertentu. Jadi tujuan ukurnya bukanlah untuk mengetahui apa yang mampu dilakukan oleh seseorang melainkan apa yang cenderung ia lakukan.

Penelitian ini akan mengkaji golongan tes yang masuk dalam klasifikasi tes prestasi. Dalam hal ini, tes prestasi belajar yang diberikan secara kelompok dalam tingkatan nasional yang biasa disebut dengan UN.

Sehubungan dengan prestasi belajar Winkel (1996:226) mengemukakan bahwa prestasi belajar merupakan bukti keberhasilan yang telah dicapai oleh seseorang. Maka prestasi belajar merupakan hasil maksimum yang dicapai oleh seseorang setelah melaksanakan usaha-usaha belajar. Prestasi belajar dapat diukur melalui tes yang sering dikenal dengan tes prestasi belajar.

Tes prestasi belajar bila dilihat dari tujuannya yaitu mengungkap keberhasilan seseorang dalam belajar. Testing pada hakikatnya menggali informasi yang dapat digunakan sebagai dasar pengambilan keputusan. Tes prestasi belajar berupa tes yang disusun secara terencana untuk mengungkap performasi maksimal subyek dalam menguasai bahan-bahan atau materi yang telah diajarkan. Dalam kegiatan pendidikan formal tes prestasi belajar dapat berbentuk ulangan harian, tes formatif, tes sumatif, bahkan ebtanas dan ujian-ujian masuk perguruan tinggi (Azwar, 2005 : 8 -9).

Bentuk tes yang digunakan dilembaga pendidikan dapat dikategorikan menjadi dua, yaitu tes objektif disini dilihat dari sistem penskorannya, siapa saja yang memeriksa lembar jawaban tes akan menghasilkan skor yang sama. Tes yang non objektif adalah yang sistem penskorannya dipengaruhi oleh pemberi skor. Dengan kata lain dapat dikatakan bahwa tes yang objektif adalah yang penskorannya objektif, sedangkan yang non objektif sistem penskorannya dipengaruhi subjektivitas pemberi skor (Mardapi, 2007:70).

## 4. Teori Tes

### a. Teori Tes Klasik ( *Classical Test Theory* )

Teori tes klasik atau disebut teori skor murni klasik (Allen & Yen, 1979:57) didasarkan pada suatu model aditif, yakni skor amatan merupakan penjumlahan dari skor sebenarnya dan skor kesalahan pengukuran. Jika dituliskan dengan pernyataan matematis, maka kalimat tersebut menjadi :

$X = T + E$

dengan :

X : skor amatan,

T : skor sebenarnya,

E : skor kesalahan pengukuran *(error score).*

Kesalahan pengukuran yang dimaksud dalam teori ini adalah kesalahan yang tidak sistematis atau acak. Kesalahan ini merupakan penyimpangan secara teoritis dari skor amatan diharapkan. Kesalahan pengukuran yang sistematis dianggap bukan merupakan kesalahan pengukuran.

Ada beberapa asumsi dalam teori tes klasik. Skor kesalahan pengukuran tidak berinteraksi dengan skor sebenarnya, merupakan asumsi yang pertama.

Asumsi yang kedua adalah skor kesalahan tidak berkorelasi dengan skor sebenarnya dan skor-skor kesalahan pada tes-tes yang lain untuk peserta tes *(testee)* yang sama. Ketiga, rata-rata dari skor kesalahan ini sama dengan nol.

Asumsi-asumsi pada teori tes klasik ini dijadikan dasar untuk mengembangkan formula-formula dalam menentukan validitas dan reliabilitas tes. Validitas dan reliabilitas pada perangkat tes digunakan untuk menentukan kualitas tes. Kriteria lain yang dapat digunakan untuk menentukan kualitas tes adalah indeks kesukaran, daya pembeda dan efektivitas distraktor.

**b. Teori Respon Butir**

Dalam evaluasi yang dilaksanakan dalam pendidikan, siswa menjawab butir soal suatu tes yang berbentuk pilihan ganda dengan benar, biasanya diberi skor 1 dan 0 jika menjawab salah. Pada penyekoran dengan pendekatan teori tes klasik, kemampuan siswa dinyatakan dengan skor total yang diperolehnya. Prosedur ini kurang memperhatikan interaksi antara setiap orang siswa dengan butir. Pendekatan teori respons butir merupakan pendekatan alternatif yang dapat digunakan dalam menganalisis suatu tes. Ada dua prinsip yang digunakan pada pendekatan ini, yakni prinsip relativitas dan prinsip probabilitas. Pada prinsip relativitas, unit dasar dari pengukuran bukanlah siswa atau butir, tetapi lebih kepada kemampuan siswa relatif terhadap butir. Jika $\beta_n$ merupakan indeks dari kemampuan siswa ke n pada trait yang diukur, dan $\vartheta_i$ merupakan indeks dari tingkat kesulitan dari butir ke-i relative yang terkait dengan kemampuan yang diukur, maka bukan $\beta_n$ atau $\vartheta_i$ yang merupakan unit pengukuran, tetapi lebih kepada perbedaan antara kemampuan dan dari siswa relatif terhadap tingkat kesulitan butir atau ($\beta_n$ - $\vartheta_i$) perlu dipertimbangkan. Sebagai alternatifnya perbandingan antara kemampuan terhadap tingkat kesulitan dapat digunakan. Jika kemampuan dari siswa melampaui tingkat kesulitan butir, maka respons siswa diharapkan benar, dan jika kemampuan siswa kurang dari tingkat kesulitan butir, maka respons siswa diharapkan salah (Keeves dan Alagumalai, 1999:24).

**c. Asumsi**

Hambleton & Swaminathan (1985: 16) dan Hambleton, Swaminathan, & Rogers (1991: 9) menyatakan bahwa ada tiga asumsi yang mendasari teori respon butir, yaitu unidimensi, independensi lokal dan invariansi parameter. Ketiga asumsi dapat dijelaskan sebagai berikut.

**d. Model Teori Respons Butir**

Teori respons butir membangun suatu model yang menghubungkan karakteristik butir dangan karakteristik peserta. Dengan sejumlah syarat tertentu, model hubungan ini dibuat untuk berlaku bebas bagi kelompok butir dan kelompok peserta mana saja yang memenuhi syarat itu. Karakteristik butir dan karakteristik peserta dihubungkan oleh model yang berbentuk fungsi atau lingkungan grafik. Sejumlah syarat yang dimaksud dinyatakan dengan sejumlah parameter. Ada parameter butir dan parameter peserta.

Ada beberapa model butir respons atau karakteristik butir, diantaranya adalah model logistik. Selanjutnya sesuai dengan batasan dan rumusan permasalahan penelitian, yang akan dibahas lebih lanjut adalah model logistik.

Model logistik terdiri dari model logistik satu-parameter (IP), model logistik dua-parameter (2P), dan model logistik tiga-parameter (3P). Ketiganya berlaku untuk butir dengan respons dikotomi, yaitu butir yang skornya benar atau salah. Sesuai dengan namanya model tiga parameter memiliki tiga parameter butir yaitu parameter tingkat kesukaran butir, daya pembeda, dan dugaan pseudo (*pseudo guessing*). Model logistik dua parameter memiliki dua parameter butir yaitu, parameter tingkat kesukaran butir, dan

 ISBN 978-602-14215-0-5

parameter daya pembeda, sedangkan parameter dugaan pseudo dianggap nol. Model logistik satu parameter memiliki satu parameter butir yaitu, parameter tingkat kesukaran, sedangkan seperti parameter daya beda dianggap sama, dan parameter dugaan pseudo sama dengan nol.

**e. Metode Pendeteksian *DIF***

Osterlind (1983), mengemukakan lima teknik untuk mendeteksi ada tidaknya *DIF* pada butir-butir tes, yaitu : (1) analisis variasi, (2) transformasi indeks kesulitan butir, (3) kai kuadrat, (4) analisis respons distraktor, dan (5) kurva karakteristik butir. Camili dan Shepard (1994), mengemukakan tujuh teknik untuk mendeteksi bias, namun demikian lima teknik diantaranya yang berdasarkan teori tes klasik tidak direkomendasikan untuk digunakan kelima teknik tersebut adalah: (1) transformasi indeks kesukaran butir, (2) penyesuaian transformasi indeks kesulitan butir, (3) prosedur aturan emas, (4) analisis variasi, dan (5) perbedaan pada korelasi biserial titik butir. Dua teknik lainya yang direkomendasikan untuk digunakan adalah: (1) kurva karakteristik butir, berdasarkan teori respons butir, dan (2) tabel kontingensi. Baik Osterlind (1983), maupun Camili dan Shepard (1994) setuju bahwa teknik kurva karakteristik adalah yang paling baik untuk mendeteksi *DIF*. Oleh karena itu, hanya teknik inilah yang akan dibahas lebih lanjut dan digunakan untuk evaluasi *DIF* dalam penelitian ini.

Pendekatan kurva karakteristik butir untuk mendeteksi dan mengoreksi butir tes *DIF* diturunkan dari respon. Pendekatan ini sangat baik dari seluruh model untuk menyisir *DIF*. Sebagai gambaran singkat, probabilitas untuk memperoleh jawaban benar pada suatu butir tes untuk setiap dua kelompok dibandingkan secara grafikal, kurva karakteristik butir menunjukan fungsi probabilitas. Kurva karakteristik butir yang dihasilkan untuk setiap dari dua kelompok harus sama untuk butir tes, dinyatakan *DIF* tidak ada. Butir tes tidak mengandung *DIF* jika seluruh individu yang memiliki kemampuan yang sama memililki probabililtas sama untuk menjawab butir tes dengan benar, tanpa memperhatikan keanggotaan sub kelompok.

Teknik kurva karakteristik butir untuk mendeteksi *DIF* adalah dengan membandingkan perbedaan pada kurva karakteristik butir dua kelompok yang diteliti. Perbedaan pada kurva karakteristik butir antara dua kelompok menunjukan bahwa pada tingkat kemampuan yang sama pengambil tes dari kelompok acuan (A) dan kelompok lainya (B) tidak memiliki probabilitas yang sama untuk menjawab butir a, b,dan c oleh karena itu perbedaan kurva karakteristik butir untuk dua kelompok secara matematis dapat disajikan sebagai perbedaan dalam parameter a, b, atau c, atau kombinasi dari ketiganya.

Perbedaan parameter tersebut menyebabkan *DIF* terjadi dalam kategori umum: pertama, *DIF* konsisten atau uniform, terjadi jika kurva karakteristik butir berbeda dan tidak saling berpoptongan/bersalingan. Hal ini terjadi kedua kurva karakteristik butir memiliki parameter a sama, jadi, berbeda hanya pada parameter b. Kedua, *DIF* tidak konsisten atau non-uniform, terjadi jika kurva karakteristik butir berbeda tetapi berpotongan pada suatu titik pada skala $\theta$. Oleh karena itu, *DIF* untuk dan pada kelompok tertentu seimbang atau sampai pada timgkat tertentu saling meniadakan satu sama lainnya. *DIF* positif mungkinseluruhnya atau sebagaian saling meniadakan tergantung pada pasangan kedua kurva karakteristik butir ( Camili dan Shepard, 1994 ).

Area antara dua kurva karakteristik butir memberikan kesan derajat *DIF*. Jika dua kurva karakteristik butir berpotongan, sebagian area dikatakan sebagai *DIF* positif dan sebagian yang lainya sebagai *DIF* negatif. Pada kasus khusus seperti ini kedua daerah antara kurva karakteristik ini dianggap sebagai *DIF* positif dan keduannya dijumlahkan membentuk indeks secara keseluruhan.

**5. Penerapaan Teori Respons Butir untuk mendeteksi *DIF***

Ada dua pendekataan berbeda, tapi saling berhubungan, untuk menguji *DIF* dengan model teori respons butir, yaitu (1) pengukuran *DIF*, dan (2) uji hipotesis statistic *DIF*.

**a. Pengukuran *DIF***

untuk pengukuran *DIF* dugunakan suatu indeks untuk menyatakan besarnya *DIF*. Ada beberapa cara yang dapat digunakan untuk mengukur ukuran perbedaan performans butir tes. Camilli dan Shepard (1994), mengemukakan empat metode, yaitu : (1) indeks area sederhana. (2) indeks perbedaan probabilitas, (3) perbedaan parameter kesulitan butir, b, dan (4) metode kurva karakteristik butir untuk sampel kecil. Sebagaiman yang sangat direkomendasikan oleh Camilli dan Sheparad (1994), untuk mengukur *DIF* pada penelitian ini mengunakan metode indeks perbedaan probabilitas.

**b. Uji hipotesis statistic *DIF***

seperti halnya dalam mengukur ukuran DIF, uji hipotesis statistic DIF dapat juga dilakukan dengan beberapa pendekatan. Camilli dan Shepard (1994), mengemukakan lima teknik, yaitu : (1) uji perbedaan parameter kesulitan butir,b, (2) metode penyimpangan butir, (3) uji khi kuadrat Lord, (4) distribusi sampling empiric indeks DIF, dan (5) mengukur perbandingan model teori respon butir. Pendekatan terakhir masih dibagi lagi menjadi empat kategori utama, yaitu : (a) umum, (b) logleniar, (c) informasi terbatas, dan (d) teori respon butir sebagaimana diterapkan pada teknik penyimpangan butir.

**C. Metode Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif. Pendekatan kuantitatif dilaksanakan dengan ex-post facto. karenanya dalam penelitian ini tidak dilakukan penelitian apapun terhadap variabel penelitian. Data utama yang digunakan dalam penelitian ini berupa jawaban peserta UN Matematika SMA paket utama 3 di Jawa Tengah tahun pelajaran 2011/2012.

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.

**1. Populasi dan Sampel**

Wilayah propinsi Jawa Tengah dibagi menjadi 35 Daerah Tingkat II yang terdiri dari 6 dan 29 kabupaten. Dalam pelaksanaan UN SMA di propinsi Jawa Tengah, untuk setiap daerah tersebut dijadikan rayon penyelenggara UN SMA, sehingga pada pelaksanaan UN SMA tahun pelajaran 2011/2012 terdapat 35 rayon. Sedangkan Populasi yang diambil dalam penelitian ini adalah respon siswa terhadap UN Matematika untuk 3 kabupaten dan 3 kotamadya.

Mata pelajaran SMA yang di UN kan tahun pelajaran 2011/2012 dipropinsi Jawa Tengah untuk program Ilmu Pengetahuan Alam (IPA) adalah : (1) Bahasa Indonesia, (2) Bahasa Inggris, (3) Matematika, (4) Fisika (5) Kimia (6) Biologi. Namun dalam populasi penelitian ini adalah mata pelajaran Matematika.

Pengambilan sampel ditempuh melalui dua tahap. Tahap pertama untuk menentukan wilayah, tahap kedua untuk menentukan jawaban peserta tes UN Matematika SMA paket soal utama 3 tahun pelajaran 2011/2012. Pengambilan sampel tahap pertama ditempuh berdasarkan pertimbangan tertentu atau sampel purposif. Pertimbangan tersebut adalah dengan memperhatikan perbedaan karakteristik wilayah penyelenggara UN.

2. **Variabel Penelitian**


ISBN 978-602-14215-0-5

Variabel penelitian dalam hal ini respon siswa peserta tes UN Matematika SMA tahun pelajaran 2011/2012 di Jawa Tengah dari 3 kabupaten dan 3 kota madya. Sehingga variabel dalam penelitian ini adalah seluruh peserta UN matematika.

**3. Teknik Pengumpulan Data**

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan teknik dokumentasi, yaitu mengutip respons siswa pada perangkat tes UN Matematika SMA paket utama 3 di jawa Jengah tahun 2011/2012.

Data-data tersebut dapat diperoleh dari Kantor Wilayah Departemen Pendidikan dan Kebudayaan Jawa Tengah di Semarang.

**4. Teknik Analisis Data**

Analisis terhadap karakteristik butir tes UN dilakukan melalui dua pendekatan. Melalui pendekatan teori tes klasik, secara empiris mutu butir-butir ditentukan oleh satatistik butir tes yang meliputi tingkat kesukaran p, daya pembeda d, dan efektifitas pengecoh. Kualitas tes ditentukan oleh statistik tes yang diantaranya meliputi rata-rata skor tes. Variansi dan simpangan baku skor tes, reliabilitas tes, dan kesalahan baku pengukuran. Statistik butir dan statistik tes diperoleh dari program *Iteman* versi 3,0.

Selain dengan pendekatan teori tes klasik, karakteristik butir tes dianalisis juga dengan menggunakan pendekatan teori respons butir model logistik tiga parameter. Butir-butir tes yang berdasarkan analisis kuantitatif klasik masuk dalam kategori tidak baik, tidak disertakan pada analisis ini. Secara empiris, kualitas butir ditelaah berdasarkan kecocokan data dengan model dan nilai parameter butir.kecocokan butir dengan model dan nilai parameter butir. Kecocokan butir drngan modeln dapat dilihat dari nilai distribusi chi-kuadrat hasil penaksiran ( $X^2$ hit ) dan derajat kebebasan ( df ) masing-masing butir tes. Suatu butir tes dikatakan cocok dengan model apabila nilai $X^2$ hit lebih kecil dari nilai kritik $X^2$ tabel sesuai dengan df butir tes yang bersangkutan pada taraf kepercayaan $\alpha$ = 0,05. Parameter butir meliputi daya beda (a), tingkat kesukaran (b), faktor tebakan (c), dan kecocokan data dengan model ditaksir dengan menggunakan program *billog* versi 3,07 dari Bock dan Mislevy, (1990). Fungsi informasi butir dihitung dengan program komputer Quartro Profersi 8.0 pada skala kemampuan $\theta$ antara -3,0 dan 3,0 dengan interval 0,5.

Pendektesian *DIF* dilakukan terhadap seluruh butir yang berdasarkan analisis klasik masuk dalam katagori ‘’baik’’. Oleh karena itu, sebelumnya dicari setatisik butir dari setiap butir perangkat tes menurut pendekatan teori tes klasik dengan program *iteman.* Penyetaraan parameter butir dikerjakan dengan program *Bilog* versi 3.07 dari bockdan mislevy (1990) dengan metode tes jangkar. Pendektesian *DIF* dilakukan dalam dua tahap. Pendeteksian tahap pertama dilakukan terhadap setiap butir tes dan sebagai tes jangkar adalah butir-butir tes lainnya.

Hampir sama dengan tahap ke pertama, pendeteksian tahap kedua juga dilakukan terhadap setiap butir tes tetapi butir-butir tes yang pada pendeteksian *DIF* tahap pertama secara stastisik menunjukan adanya *DIF* dikeluarkan dari tes jangkar. Uji stastistik kembali dilakukan terhadap setiap butir tes dengan cara dan taraf signifikasi yang sama dengan tahap pertama.

Seperti halnya pada penelitian penaksiran parameter butir, untuk mendeteksi DIF juga di perlukan dua langkah pokok dengan sedikit modifikasi pada desain file data masukan.

**Daftar Pustaka**

Camili, Gregory & Shepard, Lorrie A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publication.

Naga, Dali S. (1992). *Pengantar Teori Skor Pada Pengukuran Pendidikan*. Jakarta: Gunadarma.

Ahmann, J.S & Glock, M.D. (1981). *Evaluating Student progress: Prinsiples of testa and measurements.* Boston: Allyn and Bacon.

Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Belmon, california: Woodsworth, Inc.

Ridho, Ali. (2009). *Bias Gender Dalam Tes.* UIN-Malang Press: Malang

Saifuddin Azwar. (1997). *Reliabilitas dan validitas (edisi ke-3).* Yogyakarta: Pustaka Pelajar.

Cronbach, L.J. (1970). *Essential of psychological testing*, New York: Harper and Row publisher.

Ebel, Robert L. (1972). *Essentials of educational measurement.* Englewood Cliffs, New Jersey: Prentice Hall Inc.

Gronlund, Norman E. (1981). *Measurement and evaluation in teaching.* New York: Macmillan.

Hambleton, R.K. (1989). Principles and selected applications of item response Thory. Dalam R.L. Linn(Ed). *Education measurement hal. 147-200* New York: Macmillan.

Hambleton, R.K. & Swaminathan, H. (1985). *Ite*m response theory. Boston: Kluwer.

Mehrens, W.A & Lehmann, I.J. (1973). *Measurement and evalution in education and psycology. New York: Holt, Rinehart, and Winston, Inc.*

Woethen, B.R & Sanders, J. R. (1973). *Educational evalution; Theory and Practice.* Belmont, California: Wwodswort Publishing Company, Inc.

Hullin, C. L., et al. (1983). *Item response theory : Application to psichologycal measurement.* Homewood, IL : Dow Jones-Irwin.


ISBN 978-602-14215-0-5

# LEMBAR TANYA JAWAB

SEMINAR NASIONAL EVALUASI PENDIDIKAN (SNEP) 1
PPs UNNES, 13 JULI 2013

Ruang : 04
Moderator : Dr. Edi UTOMO

Nama Penyaji : Samsul Hadi
Instansi : UPS Tegal
Judul : Keberfungsian Item Differensial pada Perangkat tes UN Matematika

Nama Peserta : Hadromi
Instansi :
Pertanyanaan

1 Gambaran umum thg hasil akhir !

Jawab

1 Dari 40 soal MTK nti akan dianalisis karakteristik menggunakan Software Iteman.
Dif ini yg nantinya akan mempengaruhi kemampuan atau biasnya yang jauh lebih besar pengaruhnya

Pemakalah