

KEBERFUNGSIAN BUTIR DIFERENSIAL PERANGKAT TES EBTANAS KIMIA SEKOLAH MENENGAH UMUM DI JAWA TENGAH

Oleh:

Purwo Susongko dan Djemari Mardapi

Abstrak

Penelitian ini bertujuan untuk mengetahui: (1) karakteristik butir dari perangkat tes Kimia pada Ebtanas SMU di Jawa Tengah tahun pelajaran 1998/1999 ditinjau dari teori respons butir, (2) keberfungsian butir diferensial yang terjadi antara peserta tes dari rayon-rayon nonkodya dan peserta tes dari rayon kodya.

Populasi penelitian ini adalah seluruh lembar jawaban peserta tes yang menggunakan perangkat tes Kimia pada Ebtanas paket utama I dengan kode U1/D15/99 di Jawa Tengah tahun Pelajaran 1998/1999 yang berjumlah 17.134 lembar jawab berasal dari 6 rayon kodya dan 3 rayon nonkodya. Pengambilan sampel ditempuh melalui dua tahap. Tahap pertama untuk menentukan wilayah, tahap kedua untuk menentukan jawaban peserta tes Ebtanas Kimia SMU paket utama I (U1/D15/99) tahun pelajaran 1998/1999. Hasil analisis tes dengan *DIF* menunjukkan:

- 1) Tujuh butir menguntungkan peserta tes dari rayon Kodya Surakarta, sedangkan 3 butir yang lain menguntungkan peserta tes dari rayon-rayon nonkodya.
- 2) Tiga butir menguntungkan peserta tes dari rayon Kodya Salatiga, sedangkan 5 butir yang lain menguntungkan peserta tes dari rayon-rayon nonkodya.
- 3) Empat butir menguntungkan peserta tes dari rayon Kodya Semarang, sedangkan 2 butir yang lain menguntungkan peserta tes dari rayon-rayon nonkodya.
- 4) Enam butir menguntungkan peserta tes dari rayon Kodya Pekalongan, sedangkan 6 butir yang lain menguntungkan peserta tes dari rayon-rayon nonkodya.
- 5) Tak satupun dari butir-butir tes yang menguntungkan peserta tes dari rayon Kodya Tegal.

Kata kunci: butir diferensial, perangkat tes, ebtanas.

Pendahuluan

Salah satu aspek dalam penyelenggaraan pendidikan nasional adalah penilaian hasil belajar siswa. Penyelenggaraan evaluasi pendidikan secara nasional di Indonesia telah diatur melalui UU No. 2 Tahun 1989. Undang-Undang tersebut lebih lanjut menyatakan bahwa: "Dalam rangka pembinaan satuan pendidikan, pemerintah melakukan penilaian setiap satuan pendidikan secara berkala" (pasal 46 (1) UU No. 2 tahun 1989). Sebagai pelaksanaan Undang-undang sistem pendidikan nasional tersebut pemerintah menye-

lenggarakan evaluasi belajar tahap akhir nasional (Ebtanas) pada jenjang pendidikan dasar dan menengah.

Tujuan Ebtanas menurut Depdikbud (1999) adalah: (1) untuk memperoleh informasi tentang mutu hasil pendidikan secara nasional. (2) mengukur pencapaian hasil belajar siswa baik sekolah/madrasah negeri maupun sekolah/madrasah swasta. (3) memperoleh gambaran perbandingan mutu pendidikan pada sekolah/madrasah, antarsekolah/madrasah, dan antarwilayah dari tahun ke tahun; (4) menjadi bahan penentuan kebijakan pembinaan sekolah/madrasah; dan (5) sebagai bahan pertimbangan dalam memberikan Surat Tanda Tamat Belajar dan seleksi masuk ke jenjang pendidikan yang lebih tinggi.

Hasil evaluasi akhir biasanya diberikan dalam bentuk nilai Ebtanas murni (NEM). Penggunaan nilai Ebtanas murni (NEM) disesuaikan dengan tujuan Ebtanas sehingga NEM diharapkan dapat digunakan untuk melihat gambaran kemampuan siswa secara murni dan sebagai indikator keberhasilan pelaksanaan sistem pembelajaran. Demikian pula NEM digunakan untuk membandingkan mutu pendidikan antarsekolah/madrasah dan antarwilayah dari tahun ke tahun. Bagi siswa, NEM digunakan sebagai pertimbangan kelulusan dan sebagai alat seleksi masuk ke jenjang pendidikan yang lebih tinggi. Berbeda dengan penggunaan NEM di sekolah dasar (SD) dan sekolah lanjutan tingkat pertama (SLTP), NEM di sekolah menengah umum (SMU) tidak digunakan untuk seleksi masuk ke jenjang pendidikan yang lebih tinggi. Namun demikian, NEM SMU tetap digunakan sebagai pertimbangan kelulusan siswa di sekolah menengah umum.

Penggunaan NEM sedemikian luas sehingga dapat menentukan kebijakan yang akan dilaksanakan oleh penyelenggara pendidikan dan menentukan kelulusan serta keberhasilan siswa dalam seleksi masuk ke jenjang pendidikan yang lebih tinggi. Oleh karena itu, perangkat tes (soal) Ebtanas harus benar-benar dapat mengukur apa yang seharusnya diukur dan memberikan hasil pengukuran yang dapat dipercaya. Untuk mencapai keadaan tersebut tingkat kesukaran butir diupayakan mendekati atau sama dengan kemampuan peserta tes, daya beda butir diusahakan setinggi mungkin, dan faktor tebakan diupayakan serendah mungkin sehingga memberikan fungsi informasi butir yang maksimal. Penyusunan butir-butir tes Ebtanas selama ini telah dilakukan berdasarkan penelaahan butir tes secara kualitatif yang meliputi telaah materi, konstruksi, dan bahasa. Secara teoretis peringkat tes Ebtanas dapat dikatakan

sudah baik karena telah dianalisis secara kualitatif yaitu melalui penilaian subjektif dari beberapa ahli atau guru dalam bidang studi sebelum perangkat tes tersebut diujikan. Namun, secara empiris perangkat tes tersebut, khususnya pada perangkat Ebtanas SMU belum dapat dikatakan demikian, karena belum pernah diujicobakan dan dianalisis secara kuantitatif (Darsana, 1996 dan Safari, 1995). Oleh karena itu, dapat saja terjadi bahwa tes atau alat ukur yang dihasilkan masih mengandung beberapa kelemahan.

Berkaitan dengan karakteristik butir, penggunaan NEM akan sesat atau menyebabkan masalah jika ternyata butir-butir tes Ebtanas memihak kelompok tertentu sehingga dua peserta ujian dari kelompok yang berbeda dengan kemampuan yang sama mempunyai peluang menjawab benar yang berbeda. Kondisi demikian rupa dapat merupakan petunjuk adanya bias butir, dan bila bias butir ini terjadi merupakan ancaman terhadap validitas konstruk perangkat tes (Camilli dan Shepard, 1994). Tes dengan validitas yang rendah menyebabkan tes tersebut tidak dapat mengukur kemampuan siswa secara objektif. Berkaitan dengan objektivitas tes, Gronlund (1981) menyatakan bahwa tes yang disusun dengan baik akan dapat menggambarkan kemajuan belajar siswa secara objektif dan tanpa memihak kelompok tertentu sehingga hasilnya dapat dipercaya.

Prosedur awal untuk mengetahui terdapat tidaknya bias butir pada suatu butir tes dilakukan dengan analisis *Differential Item Functioning* (keberfungsian butir diferensial) yang selanjutnya disingkat *DIF*. Analisis *DIF* digunakan untuk mengidentifikasi butir yang memiliki perbedaan fungsi untuk kelompok siswa yang berbeda. Suatu butir tes menunjukkan *DIF* jika siswa-siswa yang mempunyai kemampuan yang sama, tetapi berasal dari kelompok yang berbeda, tidak 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 butir-butir tampak relatif lebih sulit untuk satu kelompok dibandingkan dengan kelompok lain. Hanya jika suatu butir relatif lebih sulit untuk satu kelompok dan kesulitan tersebut tidak relevan terhadap konstruk tes maka butir tes tersebut dikatakan bias. Dengan demikian, suatu butir tes yang mengandung *DIF* tidak otomatis butir tersebut bias karena masih banyak prosedur lain yang digunakan untuk menentukan bias atau tidaknya suatu butir tes termasuk analisis logik dari para ahli bidang studi.

Hasil rata-rata NEM selama empat tahun dari tahun pelajaran 1993/1994 sampai dengan 1996/1997 secara nasional tergolong rendah yaitu di bawah 6. Dalam empat tahun pelajaran tersebut (1993/1994 - 1996/1997) mata pelajaran yang mempunyai peringkat rata-rata NEM terendah adalah Kimia disusul kemudian oleh Fisika dan urutan ketiga ditempati Matematika (Mardapi dkk., 1999). Hasil rata-rata NEM tersebut menunjukkan bahwa tiga mata pelajaran yang mempunyai peringkat terendah adalah mata pelajaran-mata pelajaran yang merupakan ilmu-ilmu dasar bagi pengembangan ilmu pengetahuan dan teknologi. Apabila hal ini terus berlangsung merupakan suatu pertanda yang tidak baik dalam pencapaian prioritas pembangunan pendidikan nasional. Hal ini disebabkan prioritas pembangunan pendidikan nasional sampai saat ini adalah peningkatan pendidikan dalam rangka penguasaan ilmu pengetahuan dan teknologi (Iptek). Oleh sebab itu, pada jenjang pendidikan menengah, pendidikan Iptek diarahkan pada peningkatan penguasaan ilmu-ilmu dasar Iptek seperti Matematika, Bahasa dan Ilmu Pengetahuan Alam (Depdikbud, 1996).

Kimia sebagai salah satu ilmu dasar bagi pengembangan Iptek yang diajarkan di sekolah menengah umum ternyata belum dapat dikuasai siswa dengan baik. Hal ini dapat ditunjukkan bahwa selama tahun pelajaran 1993/1994 hingga 1996/1997 perolehan rata-rata NEM Kimia secara nasional belum mencapai 5 (Mardapi dkk, 1999). Rendahnya rata-rata NEM kimia tersebut menjadi suatu fenomena yang menarik untuk diteliti terutama berkaitan dengan faktor-faktor yang dapat mempengaruhi rendahnya perolehan rata-rata NEM. Faktor-faktor tersebut di antaranya kemampuan guru, kemampuan siswa, sarana dan prasarana pendidikan, mutu butir tes, dan tingkat kesukaran butir (Safari, 1996). Berkaitan dengan mutu butir tes, selain optimalisasi parameter-parameter butir untuk mendapatkan fungsi informasi butir yang maksimal perlu diperhatikan pula objektivitas dari butir-butir tes tersebut. Hal ini menjadi sangat penting karena penggunaan NEM yang begitu luas dan menimbulkan masalah jika ternyata butir-butir tes Ebtanas mengandung bias. Seperti telah dikemukakan pada alinea-alinea terdahulu, untuk mengetahui ada tidaknya bias butir dilakukan analisis *DIF*.

Penyelenggaraan Ebtanas di Jawa Tengah dilaksanakan oleh 35 penyelenggara atau rayon berdasarkan wilayah kerja Propinsi Jawa Tengah. Rayon-rayon tersebut meliputi 29 kabupaten dan 6 kodya (Depdikbud, 1999). 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 maupun politik. Perbedaan tersebut menyebabkan latar belakang siswa yang mengikuti Ebtanas pada tiap rayon penyelenggara Ebtanas tidak sama. Keadaan demikian sangat potensial bagi timbulnya *DIF* pada butir-butir tes Ebtanas yang disebabkan perbedaan latar belakang siswa pada tiap-tiap rayon penyelenggara Ebtanas. Sebagaimana dinyatakan oleh Madaus, *et al.* (Sudarsono, 1984) bahwa latar belakang siswa atau keluarga (dalam hal ini status sosial ekonomi) berpengaruh pada pencapaian prestasi belajar siswa.

Rata-rata NEM Kimia selama dua tahun terakhir (1997/1998-1998/1999) di Jawa Tengah hanya mencapai di bawah 4,5 (Depdikbud, 1999). Hanya lima rayon yang memperoleh rata-rata NEM Kimia di atas 5 pada Ebtanas tahun 1998/1999. Dari lima rayon itu, tiga di antaranya rayon-rayon kodya yaitu Kodya Magelang, Surakarta, dan Semarang. Untuk Kodya Magelang bahkan mencapai rata-rata NEM Kimia sebesar 5.77. Besarnya perolehan rata-rata NEM yang tinggi pada sebagian rayon kodya bila dibandingkan dengan perolehan rata-rata NEM rayon-rayon kabupaten dapat menimbulkan masalah. Masalah itu adalah apakah perbedaan tersebut disebabkan perbedaan kemampuan atau karena adanya bias butir yang menguntungkan bagi siswa yang berada pada rayon-rayon kodya. Terjadinya bias butir ini sangat dimungkinkan karena latar belakang siswa pada rayon-rayon kodya yang relatif merupakan daerah perkotaan akan berbeda dengan latar belakang siswa pada rayon-rayon kabupaten yang lebih banyak merupakan daerah pedesaan.

Untuk mengetahui apakah terjadi bias butir pada butir-butir tes Ebtanas Kimia di Jawa Tengah tahun 1998/1999 maka perlu dilakukan penelitian *DIF* yang didasarkan pada perbedaan wilayah penyelenggara Ebtanas.

Pembahasan

Analisis Butir dengan program *Bilog* versi 3.07 didahului dengan analisis klasik dengan program *Iteman* versi 3.0. Dari hasil analisis klasik, diketahui bahwa 53 butir tes dari 60 butir tes yang dianalisis mempunyai nilai *r* biserial (korelasi biserial antara skor butir dan skor total) lebih besar atau sama dengan 0.30. Adapun 7 butir memiliki nilai *r* biserial berturut-turut 0.248, 0.185, 0.192, 0.172, 0.182, 0.059, dan 0.261. Hal ini menandakan butir tes cukup baik dilihat dari besarnya daya beda butir.

Demikian pula bila dilihat dari tingkat kesukaran butir, ada 55 butir dari 60 butir tes yang dianalisis mempunyai nilai tingkat kesukaran butir antara 0.3 sampai dengan 0.7. Hanya 5 butir yang mempunyai nilai tingkat kesukaran butir yang kurang baik yaitu sebesar: 0.9, 0.992, 0.950, 0.829, 0.845.

Analisis dengan teori respons butir dalam hal ini dengan program *Bilog* versi 3.07, parameter butir-butir tes diuji kesesuaiannya dengan model. Parameter-parameter tersebut adalah kecocokan butir dengan model yang secara kuantitatif dilihat dari harga χ^2 , tingkat kesukaran butir, daya beda butir, faktor tebakan dan kecocokan butir dengan model yang dipilih.

Suatu butir tergolong tidak boleh digunakan apabila butir tersebut tidak cocok dengan model yang digunakan. Kecocokan butir dengan model ditunjukkan bila nilai χ^2 dari tiap butir tidak melebihi nilai kritik χ^2 pada taraf signifikansi tertentu dan derajat kebebasan yang berbeda-beda untuk tiap butir.

Pada taraf signifikansi 0.01, hasil analisis program *Bilog* versi 3.07 menunjukkan 8 butir tidak cocok dengan model yang digunakan. Butir-butir tersebut adalah butir nomor: 2, 3, 4, 6, 10, 16, 29, dan 44. Khususnya untuk butir nomor 2, 3, 4, 6, dan 10 mempunyai nilai χ^2 yang sangat besar dan jauh dibandingkan dengan nilai χ^2 dari butir-butir yang lain. Hal ini tentunya menunjukkan bahwa butir-butir tersebut benar-benar tidak cocok dengan model yang dipilih karena makin besar beda nilai χ^2 dari nilai kritik χ^2 , semakin tidak cocok butir tersebut. Model untuk butir nomor 16, 29, dan 44 peneliti berhati-hati untuk memutuskan bahwa butir-butir tersebut tidak cocok dengan model. Hal ini disebabkan adanya perbedaan yang kecil antara nilai χ^2 tiap butir tersebut dengan nilai kritik χ^2 . Seperti halnya dijelaskan oleh Naga (1992) bahwa pencocokan model seperti ini mengandung kelemahan. Pada ukuran sampel peserta yang kecil, semua butir cenderung untuk diterima, sedangkan pada ukuran sampel yang besar, banyak butir yang cenderung untuk ditolak. Oleh karena itu, keputusan perlu dilakukan secara relatif yakni hanya memeriksa kembali butir yang memiliki χ^2 yang jauh lebih besar dari yang dimiliki oleh butir-butir lainnya. Namun demikian, peneliti memutuskan bahwa pada taraf signifikansi 0.01 butir nomor 16, 29, dan 44 termasuk butir yang tidak cocok dengan Model.

Ada 7 butir tes yang secara statistik signifikan terdapat *DIF* lima butir tes di antaranya menguntungkan peserta dari rayon Kodya Magelang yaitu butir tes nomor 2, 12, 46, 53, dan 56. Hal ini ditandai dengan nilai *SPD*- θ negatif. Artinya, dengan kemampuan yang sama, peserta dari rayon kodya Magelang

memiliki probabilitas lebih besar dibandingkan dengan peserta dari rayon-rayon nonkodya untuk menjawab dengan benar butir nomor 2, 12, 46, 53 dan 56. Perbedaan probabilitas ini masing-masing sebesar 3.21 % untuk butir nomor 2, 3, 78% untuk butir nomor 12, 8, 95% untuk butir nomor 46, 14.95% untuk butir nomor 53 dan 11% untuk butir nomor 56. Perbedaan antara *SPD- θ* dan *UPD- θ* untuk lima butir tersebut relatif kecil atau sekitar 2% sehingga *DIF* pada butir-butir tersebut konsisten, artinya berlaku sama untuk setiap harga θ atau tidak ada perpotongan kurva karakteristik butir.

Dua butir tes yang lain menguntungkan peserta dari rayon-rayon nonkodya yaitu butir tes nomor 27 dan 48. Perbedaan probabilitas ini masing-masing sebesar 11.31% untuk butir nomor 27 dan 9.81% untuk butir nomor 48. Perbedaan antara *SPD- θ* dan *UPD- θ* pada kedua butir tersebut relatif kecil atau sekitar 2% sehingga *DIF* pada kedua butir tersebut konsisten, artinya sama untuk setiap harga θ .

Ada 10 butir tes yang secara statistik signifikan terdapat *DIF*. Tujuh butir tes di antaranya menguntungkan peserta dari rayon Kodya Surakarta yaitu butir tes nomor 17, 20, 21, 28 34, 37 dan 45. Hal ini ditandai dengan nilai *SPD- θ* negatif. Artinya, dengan kemampuan yang sama, peserta dari rayon Kodya Surakarta memiliki probabilitas lebih besar dibandingkan dengan peserta dari rayon-rayon nonkodya untuk menjawab dengan benar butir nomor 17, 20, 21, 28, 34, 37 dan 45. Perbedaan probabilitas ini masing-masing sebesar 9.77% untuk butir nomor 17, 7.11% untuk butir nomor 20, 9.21% untuk butir nomor 21, 7.07% untuk butir nomor 28, 8.94% untuk butir 34, 9.58% untuk butir nomor 37 dan 3.82% untuk butir nomor 45. Kecuali butir nomor 20. Perbedaan antara *SPD- θ* dan *UPD- θ* untuk butir-butir yang lain relatif kecil atau sekitar 2% sehingga *DIF* pada butir-butir tersebut konsisten, artinya berlaku sama untuk setiap harga θ atau tidak ada perpotongan kurva karakteristik butir. Untuk butir tes nomor 20 mempunyai perbedaan *SPD- θ* dan *UPD- θ* cukup besar yaitu sekitar 7% sehingga butir tersebut dipastikan tidak konsisten artinya besarnya perbedaan probabilitas menjawab benar tidak sama untuk setiap harga θ yang berbeda, atau ada perpotongan kurva karakteristik butir.

Tiga butir tes yang lain menguntungkan peserta dari rayon-rayon nonkodya yaitu butir nomor 13, 23, dan 50. Perbedaan probabilitas ini masing-masing sebesar 9.0% untuk butir nomor 13, 7.0% untuk butir nomor 23 dan 8.09% untuk butir nomor 50. Perbedaan antara *SPD- θ* dan *UPD- θ* pada ketiga

butir tersebut relatif kecil atau sekitar 2% sehingga *DIF* pada ketiga butir tersebut konsisten, artinya berlaku untuk setiap harga θ .

Ada 7 butir tes yang secara statistik signifikan terdapat *DIF*. Tiga butir tes di antaranya menguntungkan peserta dari rayon Kodya Salatiga yaitu butir tes nomor 20 dan 32. Hal ini ditandai dengan nilai *SPD*- θ negatif. Artinya, dengan kemampuan yang sama, peserta dari rayon Kodya Salatiga memiliki probabilitas lebih besar dibandingkan dengan peserta dari rayon-rayon nonkodya untuk menjawab dengan benar butir nomor 20, 32, dan 58. Perbedaan probabilitas ini masing-masing sebesar 11.61% untuk butir nomor 20, 0.16% untuk butir nomor 32 dan 5; 68% untuk butir nomor 58. Dua butir masing-masing nomor 20 dan 58 memiliki perbedaan *SPD*- θ dan *UPD*- θ relatif kecil atau sekitar 2% sehingga *DIF* pada butir-butir tersebut konsisten. Untuk butir tes nomor 32 mempunyai perbedaan *SPD*- θ dan *UPD*- θ cukup besar yaitu sekitar 7.58% sehingga butir tersebut dipastikan tidak konsisten, artinya besarnya perbedaan probabilitas menjawab benar tidak sama untuk setiap harga θ yang berbeda.

Empat butir tes yang lain menguntungkan peserta dari rayon-rayon nonkodya yaitu butir nomor 23, 35, 43 dan 50. Perbedaan probabilitas ini masing-masing sebesar 11.56% untuk butir nomor 23, 8.28% untuk butir nomor 35, 9.08% untuk butir nomor 43 dan 9.92% untuk butir nomor 50. Perbedaan antara *SPD*- θ dan *UPD*- θ untuk butir nomor 23, 43 dan 50 relatif kecil sehingga *DIF* yang terjadi konsisten atau berlaku sama untuk setiap harga θ . Butir nomor 35 mempunyai perbedaan *SPD*- θ dan *UPD*- θ cukup besar yaitu 3.91% sehingga *DIF* yang terjadi tidak konsisten.

Ada 6 butir tes yang secara statistik signifikan terdapat *DIF* empat butir tes di antaranya menguntungkan peserta dari rayon Kodya Semarang yaitu butir tes nomor 20, 32, 33, dan 53. Hal ini ditandai dengan nilai *SPD*- θ negatif. Artinya, dengan kemampuan yang sama, peserta dari rayon Kodya Semarang memiliki probabilitas lebih besar dibandingkan peserta dari rayon-rayon nonkodya untuk menjawab dengan benar butir nomor 20, 32, 33, dan 53. Perbedaan probabilitas ini masing-masing sebesar 7.36% untuk butir nomor 20, 10, 30% untuk butir nomor 32 dan 5; 60% untuk butir nomor 33, dan 7; 29% untuk butir nomor 53. Perbedaan antara *SPD*- θ dan *UPD*- θ untuk empat butir tersebut relatif kecil atau sekitar 2% sehingga *DIF* pada butir-butir

tersebut konsisten, artinya berlaku sama untuk setiap harga θ atau tidak ada perpotongan kurva karakteristik butir.

Dua butir tes yang lain menguntungkan peserta dari rayon-rayon nonkodya yaitu butir tes nomor 14 dan 15. Perbedaan probabilitas ini masing-masing sebesar 8,34% untuk butir nomor 14 dan 10,68% untuk butir nomor 15. Perbedaan antara $SPD-\theta$ dan $UPD-\theta$ pada kedua butir tersebut relatif kecil atau sekitar 2% sehingga DIF pada kedua butir tersebut konsisten, artinya sama untuk setiap harga θ .

Ada 12 butir tes yang secara statistik signifikan terdapat DIF . Enam butir tes di antaranya menguntungkan peserta dari rayon Kodya Pekalongan yaitu butir tes nomor 4, 17, 31, 32, 34 dan 52. Hal ini ditandai dengan nilai $SPD-\theta$ negatif. Artinya, dengan kemampuan yang sama, peserta dari rayon Kodya Pekalongan memiliki probabilitas lebih besar dibandingkan dengan peserta dari rayon-rayon nonkodya untuk menjawab dengan benar butir nomor 4, 17, 31, 32, 34, dan 52. Perbedaan probabilitas ini masing-masing sebesar 6,79% untuk butir nomor 4 dan 5; 93% untuk butir nomor 17 dan 9; 06% untuk butir nomor 31 dan 2; 61% untuk butir nomor 32; 5.81% untuk butir 34; 6.99% untuk butir nomor 52. Kecuali butir nomor 17 dan 32, perbedaan antara $SPD-\theta$ dan $UPD-\theta$ untuk butir-butir yang lain relatif kecil atau sekitar 2% sehingga DIF pada butir-butir tersebut konsisten, artinya berlaku sama untuk setiap harga θ atau tidak ada perpotongan kurva karakteristik butir. Untuk butir tes nomor 17 dan 32 mempunyai perbedaan $SPD-\theta$ dan $UPD-\theta$ cukup besar yaitu sekitar 3% dan 6.31% sehingga butir tersebut dipastikan tidak konsisten artinya besarnya perbedaan probabilitas menjawab benar tidak sama untuk setiap harga θ yang berbeda, atau ada perpotongan kurva karakteristik butir.

Enam butir tes yang lain menguntungkan peserta dari rayon-rayon nonkodya yaitu butir nomor 3, 5, 12, 48, 55, dan 58. Perbedaan probabilitas ini masing-masing sebesar 9,79% untuk butir nomor 3, kemudian 6,81% untuk butir nomor 5, 9,51% untuk butir no 12, 7,20% untuk butir nomor 48, 8,69% untuk butir nomor 55, 7,80% untuk butir nomor 58. Perbedaan antara $SPD-\theta$ dan $UPD-\theta$ pada ketiga butir tersebut relatif kecil atau sekitar 2% sehingga DIF pada keenam butir tersebut konsisten, artinya berlaku untuk setiap harga θ .

Ada delapan butir tes yang secara signifikan terdapat DIF . Tidak satu pun dari delapan butir tersebut yang menguntungkan peserta dari rayon Kodya

Tegal. Butir-butir tersebut adalah butir nomor 9, 31, 50, 52, 55, 57, 58, dan 60. Dengan demikian, butir-butir tersebut semuanya menguntungkan peserta dari rayon-rayon nonkodya. Perbedaan probabilitas ini masing-masing sebesar 3,71% untuk butir nomor 9, 5,90% untuk butir nomor 31, 8,38% untuk butir nomor 50, 6,36% untuk butir nomor 52, 8,88% untuk butir nomor 55, 10,77% untuk butir nomor 57, 10,76% untuk butir nomor 58 dan 7,85% untuk butir nomor 60. Hanya ada dua butir masing-masing nomor 9 dan 58 yang mempunyai perbedaan $SPD-\theta$ dan $UPD-\theta$ relatif kecil atau sekitar 2%, sehingga DIF yang terjadi tidak konsisten. Hal ini berarti bahwa besarnya perbedaan probabilitas menjawab benar tidak sama untuk setiap harga θ yang berbeda, atau ada perpotongan untuk karakteristik kurva.

Terjadinya DIF pada suatu butir dapat menjadi petunjuk awal adanya suatu bias butir. Namun, untuk menentukan bias atau tidaknya suatu butir masih dibutuhkan penelitian lebih lanjut dan lebih mendalam. Hal ini disebabkan DIF pada suatu butir dapat disebabkan oleh kualitas proses pembelajaran, seperti halnya kemampuan guru, sarana, prasarana, dan metode pengajaran. Penyelenggaraan tes Ebtanas juga menjadi penyebab timbulnya DIF , terutama berkaitan dengan pengawasan tes dan tata tertib penyelenggaraan tes.

Kualitas proses pembelajaran dapat memberikan pengaruh yang berbeda pada dua kelompok yang mempunyai kemampuan sama. Sebagai contoh butir nomor 60. Butir ini menguji kemampuan siswa dalam menentukan bahan makanan yang mengandung ikatan peptida dengan pereaksi *biuret* dan *Xantoproteat* melalui penafsiran hasil percobaan. Bagi siswa yang tidak mendapatkan praktikum tentunya tidak dapat menjawab benar butir ini walaupun mempunyai kemampuan yang sama dengan siswa yang mendapatkan praktikum. Butir nomor 60 ini mengalami DIF yang merugikan sebesar 7,85% bagi peserta ujian rayon Kodya Tegal dibandingkan dengan peserta ujian pada rayon-rayon nonkodya. Peneliti tidak dapat memastikan apakah DIF tersebut disebabkan oleh proses pembelajaran karena untuk menyatakan sebab terjadinya DIF membutuhkan penelitian yang lebih mendalam. Namun, setidaknya penjelasan tersebut dapat membantu dalam penelitian lebih lanjut tentang terjadinya DIF pada butir ini.

Masih berkaitan dengan proses pembelajaran, butir nomor 17 menunjukkan fenomena yang menarik. Butir nomor 17 menguji kemampuan siswa untuk mendefinisikan pengertian keadaan setimbang pada reaksi kimia.

Sebenarnya kemampuan ini tidak tergolong kemampuan tingkat tinggi karena pada awal bab tentang kesetimbangan definisi ini banyak dibicarakan. Namun, ternyata butir ini mengalami *DIF* yang menguntungkan peserta ujian rayon Kodya Surakarta sebesar 9,77 % dibandingkan dengan peserta pada rayon-rayon nonkodya. Kesetimbangan kimia merupakan konsep yang abstrak sehingga untuk menanamkannya pada siswa yang dibutuhkan keterampilan tersendiri bagi guru. Sebagian guru memilih metode ceramah bermakna untuk menanamkan konsep kesetimbangan kimia pada siswa, sedangkan sebagian guru yang lain menggunakan metode demonstrasi atau percobaan di laboratorium. Dengan metode ceramah bermakna banyak sekali konsep-konsep yang telah diterima siswa dapat mengganggu pembentukan konsep yang diharapkan. Berbeda halnya bila penanaman konsep dilakukan dengan kegiatan demonstrasi atau percobaan di laboratorium, siswa akan memahami konsep sebagai sesuatu yang nyata berkesan sehingga sulit hilang dari struktur kognitif siswa.

Peneliti tidak dapat memastikan apakah keadaan demikian yang terjadi pada siswa di rayon Kodya Surakarta sehingga peserta ujian dari rayon ini dengan kemampuan sama dengan peserta ujian dari rayon-rayon nonkodya mempunyai peluang menjawab benar yang lebih besar. Hal ini disebabkan masih membutuhkan penelitian lebih lanjut untuk mengetahui sebab-sebab terjadinya *DIF* pada suatu butir.

Adanya *DIF* pada suatu butir dapat pula disebabkan oleh pengawasan dan tata tertib tes yang kadang-kadang diberlakukan berbeda untuk setiap rayon. Pengawasan yang kurang ketat akan menyebabkan peserta ujian lebih leluasa untuk bekerjasama dengan sesama peserta ujian yang lain. Karena banyak peserta yang bekerjasama dengan peserta lain, hal itu akan menghasilkan estimasi parameter yang berbeda dengan harga parameter butir yang sesungguhnya. Demikian pula apabila tata tertib tes tidak diberlakukan sama di seluruh rayon penyelenggara Ebtanas. Hal ini misalnya penggunaan kalkulator sebagai alat bantu hitung, walaupun dalam naskah butir tes dinyatakan tidak boleh menggunakan kalkulator, ternyata ada rayon-rayon tertentu yang membolehkan peserta ujian menggunakan kalkulator. Keadaan demikian tentunya merugikan peserta ujian di rayon-rayon lain, yang pada akhirnya dapat menyebabkan *DIF* pada butir-butir tertentu.

Adanya *DIF* pada suatu butir dapat pula disebabkan oleh kemampuan guru yang berbeda. Dua kelompok siswa yang mempunyai kemampuan sama

akan memberikan respons yang berbeda bila diajar oleh guru yang mempunyai kemampuan berbeda. Hal ini dapat digunakan untuk menjelaskan fenomena yang terjadi pada butir nomor 17 yang dalam menjelaskan konsep kesetimbangan kimia, peserta tes dari rayon Kodya Surakarta diuntungkan 9,77% dari peserta tes rayon-rayon nonkodya. Keadaan demikian dapat terjadi karena kemampuan guru dalam memahami konsep kesetimbangan kimia di rayon Kodya Surakarta memang lebih tinggi dibandingkan dengan kemampuan guru di rayon-rayon nonkodya.

Ada 8 butir dari seluruh butir tes yang terindetifikasi *DIF* pada butir-butir tes Ebtanas Kimia SMU paket utama I di Jawa Tengah tahun Pelajaran 1998/1999 yang mempunyai nilai SPD lebih dari 10 %. Untuk *DIF* yang terjadi pada rayon penelitian Kodya Tegal dengan besarnya SPD masing-masing 10,77% dan 10,76% terjadi pada butir nomor 57 dan 58. Butir nomor 15 dan 32 untuk *DIF* yang terjadi pada rayon penelitian Kodya Semarang dengan besarnya SPD-Q masing-masing 10,68% dan -10,30%. Butir nomor 53 dan 56 untuk *DIF* yang terjadi pada rayon penelitian Kodya Magelang dengan besarnya SPD-Q masing-masing -14,95% dan -11,00%. Butir nomor 20 dan 23 untuk *DIF* yang terjadi pada rayon penelitian Kodya Salatiga dengan besarnya SPD-Q masing-masing -11,61% dan 11,56%.

Simpulan

Berdasarkan temuan hasil penelitian dan pembahasan dapat dikemukakan beberapa hal penting berkaitan dengan perangkat tes Ebtanas Kimia SMU Paket Utama I di Jawa Tengah tahun pelajaran 1998/1999 sebagai berikut: (1) Dengan menggunakan pendekatan teori respons butir model fungsi logistik tiga parameter, ada 8 butir dari 60 butir tes yang tidak cocok dengan model. (2) Fungsi informasi butir maksimum mempunyai nilai tertinggi pada butir nomor 40 yaitu 2,05 dan terendah pada butir nomor 30 yaitu 0,04. (3) Ada 7 butir tes yang terdeteksi mengalami *DIF* antara peserta ujian dari rayon-rayon nonkodya dan peserta ujian dari rayon Kodya Magelang. Lima butir menguntungkan peserta ujian dari rayon Kodya Magelang, sedangkan 2 butir yang lain menguntungkan peserta ujian dari rayon-rayon nonkodya. Tujuh butir menguntungkan peserta ujian dari rayon Kodya Surakarta, sedangkan tiga butir yang lain menguntungkan peserta ujian dari rayon-rayon nonkodya. Tiga butir menguntungkan peserta ujian dari rayon Kodya Salatiga, sedangkan 5 butir yang lain menguntungkan peserta ujian dari rayon-rayon nonkodya.

Empat butir menguntungkan peserta ujian dari rayon Kodya Semarang, sedangkan 2 butir yang lain menguntungkan peserta ujian dari rayon-rayon nonkodya. Enam butir di antaranya menguntungkan peserta ujian dari rayon Kodya Pekalongan, sedangkan 6 butir yang lain menguntungkan peserta ujian dari rayon-rayon nonkodya. Tidak satu pun butir-butir tersebut yang menguntungkan peserta ujian dari rayon Kodya Tegal.

Saran-saran

Berdasarkan hasil analisis tes Ebtanas Kimia di SMU Jateng disarankan:

1. perlu dikembangkan bank soal agar butir-butir soal yang digunakan untuk Ebtanas telah diketahui karakteristiknya,
2. tingkat kesulitan soal sebaiknya sesuai dengan tingkat kemampuan siswa agar diperoleh fungsi informasi yang maksimum,
3. perlu penyeragaman pelaksanaan Ebtanas dan peningkatan kualitas pembelajaran agar tidak terdapat *DIF*, dan
4. perlu penelitian lebih lanjut untuk mata ujian yang lain.

Daftar Pustaka

- Camilli, Gregory & Shepard, Lorrie A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks, CA: Sage Publication.
- Darsana. (1996). "Tinjauan singkat pelaksanaan Ebtanas SMU DKI Jakarta tahun pelajaran 1995/1996". *Buletin Pengujian dan Penilaian Pendidikan*. Oktober 1996. 16-10.
- _____ (1996). "Visi dan strategi pendidikan untuk tahun 2020: Tuntutan terhadap kualitas". Ceramah Menteri Pendidikan dan Kebudayaan RI dalam *Konvensi Nasional Pendidikan Indonesia III*. Ujung Pandang. 4-7 Maret 1996.
- _____ (1999). *Juklak Ebtana/Ebtanas Tahun Pelajaran 1998/1999*. Semarang: Kanwil Depdikbud Propinsi Jawa Tengah.
- Depdikbud. (1989). *Undang-Undang Republik Indonesia No. 2 Tahun 1989 tentang Sistem Pendidikan Nasional dan Penjelasannya*. Semarang: CV Aneka Ilmu.

- Ground, Norman E. (1981). *Measurement of Education in Teaching*. New York: Macmillan.
- Hambleton, R.K. (1989). "Principle and selected applications of item response theory". Dalam R.L. Linn (Ed.). *Education Measurement*. 147-200. New York: Macmillan.
- Mardapi, Djemari dkk. (1999). *Evaluasi Penyelenggaraan Ebtanas*. Laporan Penelitian. Tidak dipublikasikan. Yogyakarta. Lembaga Penelitian IKIP Yogyakarta.
- Naga, Dali S. (1992). *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Safari. (1995). "Nilai Ebtanas Murni dan permasalahannya". *Buletin Pengujian dan Penilaian Pendidikan*. September 1995. 5-7.
- Safari. (1996). "Bias budaya dalam soal Ebtanas". *Buletin Pengujian dan Penilaian Pendidikan*. No. 01/V/Maret 1996. 4-6.
- Sudarsono, FX. (1984). *A Comparison of Academic Achievement in Indonesian Private and Public Aelementary Schools*. Disertasi doktor tidak dipublikasikan. The Florida State University.