

**MENDETEKSI BIAS BUTIR TES MATEMATIKA DAN IPA  
BERDASARKAN GENDER**

**Oleh : Heri Retnawati ([Riwati@yahoo.com](mailto:Riwati@yahoo.com))  
FMIPA Matematika UNY  
Abstrak**

Dalam suatu perangkat tes, jika terdapat butir-butir soal yang memberikan peluang menjawab benar berbeda kepada kelompok peserta tes berdasarkan jenis kelamin (gender), maka dikatakan perangkat tes tersebut memuat butir-butir yang bias. Bias butir ini menyebabkan perangkat tes bersifat tidak adil. Padahal, salah satu azas terselenggarakannya suatu tes adalah keadilan (equity). Terkait dengan hal ini dan pentingnya peranan matematika dan ilmu pengetahuan alam, perlu dideteksi muatan bias butir dalam suatu perangkat tes. Muatan bias ini dapat dideteksi dengan teori tes klasik maupun dengan teori respons butir.

Kata Kunci : bias butir, gender, teori tes klasik, dan teori respons butir

**PENDAHULUAN**

Dalam dunia pendidikan, evaluasi memegang peranan yang penting. Evaluasi dalam pendidikan dilaksanakan untuk memperoleh informasi tentang aspek yang berkaitan dengan pendidikan. Menurut Gronlund (1981), evaluasi dalam pendidikan memiliki tujuan : a) untuk memberikan klarifikasi tentang sifat hasil pembelajaran yang telah dilaksanakan, b) memberikan informasi tentang ketercapaian tujuan jangka pendek yang telah dilaksanakan, c) memberikan masukan untuk kemajuan pembelajaran, d) memberikan informasi tentang kesulitan dalam pembelajaran dan untuk memilih pengalaman pembelajaran di masa yang akan datang. Informasi evaluasi dapat digunakan untuk membantu memutuskan a) kesesuaian dan keberlangsungan dari tujuan pembelajaran, b) kegunaan materi pembelajaran, dan c) untuk mengetahui tingkat efisiensi dan efektifitas dari strategi pengajaran (metode dan teknik belajar-mengajar) yang digunakan.

Untuk melaksanakan evaluasi pendidikan, diperlukan alat (instrumen). Alat yang digunakan untuk melakukan evaluasi, salah satunya adalah tes. Tes ini digunakan untuk mengetahui informasi tentang aspek psikologis tertentu. Menurut Cronbach (Nurkholis, 2000), tes merupakan suatu prosedur sistematis untuk mengamati dan menggambarkan satu atau lebih karakteristik seseorang

dengan suatu skala numerik atau sistem kategorik. Berdasarkan hal ini, tes memberikan informasi yang bersifat kualitatif dan kuantitatif.

Terkait dengan pelaksanaan tes yang harus tetap berasaskan objektivitas, transparansi, akuntabilitas dan tidak diskriminatif, diperlukan suatu perangkat tes seleksi yang adil, yang tidak memihak pada suatu kelompok siswa tertentu. Jika pada butir suatu tes memihak kelompok tertentu, maka tes tersebut dikatakan mengandung butir bias atau mengandung keberfungsian butir diferensial (*differential item functioning, DIF*). Adanya butir diferensial ini mengakibatkan perangkat tes tersebut bersifat diskriminatif, atau memihak pada kelompok tertentu.

Pengelompokan siswa dapat ditinjau dari berbagai segi, misalnya ras, budaya, wilayah, gender dan lain-lain. Agar dalam pelaksanaan tes bersifat adil, maka dalam penyusunan butir-butirnya, perlu digunakan butir-butir yang tidak memuat bias. Terkait dengan hal ini, diperlukan suatu butir-butir yang terbebas dari DIF, misalnya ditinjau berdasarkan gender. Untuk itu diperlukan suatu pendeteksian bias butir ini berdasarkan gender. Pada tulisan ini akan dibahas mengenai perlunya pendeteksian bias butir pada perangkat tes berdasarkan gender.

## **PEMBAHASAN**

Matematika merupakan pengetahuan yang tersusun secara hierarkis, yang harus dimiliki siswa untuk dapat mempelajari berbagai ilmu pengetahuan. Begitu pula dengan pelajaran sains (IPA), yang mempunyai peranan penting dalam penerapan teknologi. Dengan kata lain, mata pelajaran matematika memiliki peran yang strategis dalam menyongsong kemajuan ilmu pengetahuan dan teknologi. Ini berarti kemampuan matematika dan IPA yang dimiliki siswa dapat digunakan untuk memprediksikan keberhasilan siswa dalam mempelajari berbagai ilmu pengetahuan dan teknologi di masa mendatang.

Hasil evaluasi dalam pembelajaran matematika dan IPA memiliki peran penting pada pembelajaran. Hasil ini digunakan siswa untuk melanjutkan studi ke jenjang yang lebih tinggi. Jika perangkat tes yang digunakan memuat bias, ini sedikit banyak menentukan kualitas jenjang pendidikan selanjutnya yang ditempuh siswa. Jika sampai ke pendidikan tinggi, akan mempengaruhi jenis pekerjaan yang dipilihnya. Mengingat peran penting hasil evaluasi ini, perlu dilakukan pendeteksian bias butir, sehingga tes yang dilakukan bersifat adil bagi semua peserta tes dan tidak menguntungkan kelompok siswa tertentu.

Untuk menjamin pada suatu perangkat tes tidak memuat butir yang mengandung DIF, diperlukan analisis DIF pada perangkat tes, khususnya berdasarkan gender. Hal ini terkait dengan hasil penelitian yang dilakukan di Kanada oleh Gierl, Khaliq dan Boughton bahwa pada perangkat tes matematika dan sains menguntungkan siswa yang berjenis kelamin laki-laki (1999: 10).

Matematika merupakan pengetahuan yang memiliki obyek dasar yang abstrak, yang berdasarkan kebenaran konsistensi, tersusun secara hirarkis dan sesuai dengan kaidah penalaran yang logis. Menurut Depdiknas (2003), matematika dipandang sebagai sarana untuk berlatih cara berfikir dan bernalar dalam menarik kesimpulan, mengembangkan aktivitas kreatif, mengembangkan kemampuan memecahkan masalah, dan mengembangkan kemampuan menyampaikan informasi. Sehubungan dengan hal ini, matematika merupakan materi yang perlu dikuasai siswa sejak dini, karena matematika merupakan bekal untuk mempelajari berbagai ilmu, bahkan merupakan dasar untuk mempelajari ilmu.

Materi-materi dalam pelajaran matematika tersusun secara hierarkis dan konsep matematika yang satu dengan yang lain saling berkorelasi membentuk konsep baru yang lebih kompleks. Ini berarti bahwa pengetahuan matematika yang diketahui siswa sebelumnya menjadi dasar pemahaman untuk mempelajari materi selanjutnya. Mengingat matematika merupakan dasar dan bekal untuk mempelajari berbagai ilmu, juga mengingat matematika tersusun secara hierarkis, maka kemampuan matematika dalam proses seleksi memberikan sumbangan yang besar dalam memprediksi keberhasilan belajar siswa pada masa selanjutnya, baik dalam mempelajari matematika sendiri ataupun mempelajari ilmu lain secara luas. Demikian pula halnya dengan pelajaran sains, yang menyiapkan siswa untuk menguasai teknologi.

Perbedaan jenis kelamin, menurut Osterlind (1983), Camilli dan Shepard (1994), Hambleton dan Rogers (1995), serta Scheuneman dan Bleistein (1999) dapat menimbulkan bias butir. Hal ini dapat dimaklumi karena menurut Kartini (1980) ada perbedaan antara pria dan wanita. Perbedaan itu meliputi: (1) pada umumnya kemampuan intelektual wanita relatif lebih rendah jika dibandingkan pria, (2) pada hakekatnya wanita lebih bersifat heterosentris dan sosial jika dibandingkan dengan pria sehingga pekerjaan sosial seperti jururawat, guru dan pekerja sosial lebih menarik bagi wanita, sedangkan pria lebih tertarik pada pekerjaan yang membutuhkan

pikiran dan lebih objektif, dan (3) wanita lebih tertarik untuk mengisi waktu senggangnya dengan pekerjaan-pekerjaan yang ringan-ringan, sementara itu pria lebih senang beristirahat untuk mengisi waktu senggangnya.

Penelitian yang memperlihatkan adanya bias butir akibat perbedaan jenis kelamin ditunjukkan oleh penelitian Barrett (2001) dan penelitian Lee (1999). Penelitian Barrett (2001) ingin melihat apakah tes ujian masuk Divisi Bisnis University of South Australia mengandung bias gender atau memang dapat membedakan antara kelompok mampu dan tidak mampu. Dalam penelitian ini, butir tes hanya dilihat dari daya beda dan bias gender. Hasil penelitian menunjukkan bahwa hanya satu dari 30 butir tes yang daya bedanya kurang baik, dan hanya satu butir tes saja yang mengandung DIF atau dalam hal ini bias gender.

Sementara itu, Lee (1999) mengkaji tentang *differential item functioning* (DIF) berbasis gender pada test pencapaian sejarah kelas 10 yang diujikan sebagai bagian dari *National Education Longitudinal Study of 1988* (NELS:88). Dilakukan beberapa analisis DIF dengan berbagai kriteria pasangan, dan temuan didampingi dengan studi survei yang membantu memvalidasi interpretasi terhadap alasan yang mendasari adanya DIF. DIF yang sesuai dengan masing-masing gender menunjukkan stereotip peran gender tradisional: pria lebih berhasil pada butir tes “maskulin”, sedangkan wanita lebih diuntungkan oleh butir tes “feminin”. Studi surveinya menguatkan bahwa baik siswa sekolah menengah pria maupun siswa sekolah menengah wanita melihat butir-butir sebagai bertipe gender sebagaimana diprediksikan oleh kesesuaian peran gender. Temuan menunjukkan bahwa keberpihakan kepada pria dalam tes ini terbatas pada bidang isi tertentu dan tidak mewakili perbedaan dalam kecakapan keseluruhan.

Selain hal-hal di atas, men-stereotip-kan, tidak menyamakan, atau menampilkan kelompok-kelompok peserta tes secara tidak menyenangkan adalah sifat-sifat tes yang tidak baik dan harus betul-betul diperhatikan oleh pengembang dan penelaah butir tes. Tes harus bebas dari hal-hal yang menyinggung perasaan, merendahkan, dan membuat orang tidak senang.

Selain sumber internal, juga ada sumber eksternal yaitu sumber bias yang tidak melibatkan tesnya sendiri namun dihasilkan dari faktor di luar situasi pengujian yang berinteraksi dengan perbedaan individual atau kelompok untuk menghasilkan bias dalam skor tes bagi seseorang maupun kelompok. Beberapa faktor yang sering disebut sebagai sumber eksternal bias adalah (1) pengaruh dari pengalaman atau latihan mengerjakan tes tertentu, (2)

pengaruh interpersonal yang melibatkan sikap, harapan, dan dialek penguji dan cara memberikan petunjuk tes, dan sifat-sifat penguji yang memotivasi dan memberi dorongan, (3) pelaksanaan tes yang bersifat individual lawan kelompok dan pengaruh dari semangat dan disiplin kelas pada pengerjaan tes, (4) tes kecepatan dan tes biasa, (5) etnis dan jenis kelamin penguji yang cenderung berinteraksi dengan etnis dan jenis kelamin subjek; dan (6) penskoran tes yang bias karena adanya *hallo effects*.

Penelitian Powel (2002) terhadap 42 anak-anak yang terdiri dari 22 pria dan 20 wanita menunjukkan anak-anak pria lebih agresif daripada wanita. Sementara itu penelitian Polk dan Erdal (2002) yang berjudul "*Gender and sport differences in the emphasis on physical and mental aspects of athletics*" menemukan bahwa pria lebih cocok berolah raga secara tim (misal basket) sedangkan wanita lebih cocok berolah raga secara individual (misal lari).

Selain itu, banyak penelitian yang menunjukkan bahwa kemampuan verbal wanita lebih tinggi daripada pria. Mungkin saja kebiasaan yang ragu-ragu dan kurang percaya diri pada wanita-lah yang menyebabkan tampilannya kurang optimum. Hal ini mungkin dikarenakan siswi yang terbiasa mendapat perlakuan tidak adil dari gurunya sehingga dapat mengurangi percaya dirinya. Brennan yang dikutip Jones, dkk (2000) mengatakan dalam dunia pendidikan yang sudah berjalan berpuluh tahun ini, wanita masih berjuang melawan ketidaksamaan gender yang substansial di sekolah. Dengan adanya ketidaksamaan ini wanita mengalami keterbatasan produktivitas sosial dan emosional. Selain itu, pengalaman pendidikan memberikan dorongan dan kemampuan untuk pencapaian dalam sektor bisnis, finansial, dan di mana saja; oleh karena itu diskriminasi pada tingkat sekolah dapat merugikan wanita.

*The American Association of University Women* menerbitkan laporan penting terkait dengan ketidaksamaan gender di dalam kelas (Jones, dkk 2000). Laporan itu menjelaskan bahwa siswi menerima perhatian dari guru yang lebih sedikit bila dibandingkan dengan siswa. Siswi juga sering memperoleh kritikan pada pekerjaan yang telah diselesaikannya tanpa menghargai kebersihan, kerapian, dan ketepatan pengumpulan pekerjaan tersebut. Sebaliknya, siswa sering mendapat pujian untuk jawabannya yang cerdas dan kemampuan bagus. Siswi sering memulai dengan keraguan dan meragukan kemampuan mereka sendiri sehingga partisipasinya di dalam kelas rendah akibatnya percaya diri dan prestasi belajarnya juga rendah.

Di sisi lain, banyak penelitian yang menunjukkan bahwa kemampuan keruangan siswa pria lebih tinggi dari pada wanita. Penelitian Maccoby dan

Jacklin pada tahun 1974; Hyde, Fennema, Lamon pada tahun 1990; Halpern pada tahun 1992; dan Stumpf pada tahun 1995 (Nowell dan Hedges, 1998) menunjukkan bahwa skor tes kemampuan berhitung pria lebih tinggi daripada skor wanita. Sementara itu penelitian Linn dan Peterson pada tahun 1986; Stumpf dan Klieme pada tahun 1989; Masters dan Sander pada tahun 1993; Hedges dan Nowell pada tahun 1995; Voyer, Voyer, Bryden pada tahun 1995 (Nowell dan Hedges, 1998) menunjukkan bahwa rata-rata skor tes kemampuan spatial untuk pria lebih tinggi daripada skor wanita.

### **METODE UNTUK MENDETEKSI BIAS BUTIR**

Ada beberapa metode untuk mendeteksi DIF dalam suatu perangkat tes. Adams (Keeves, 1992 : 180) menyatakan bahwa metode yang dapat digunakan untuk mendeteksi DIF meliputi Analisis Faktor, pengujian daya beda butir menggunakan point-biserial dan korelasi parsial, pengujian tingkat kesukaran butir menggunakan berbagai transformasi, metode ANOVA, teori respons butir atau metode *latent traits*, pendekatan Khi-kuadrat (*Chi-square*), menggunakan model Log-linear dan metode statistika Mantel-Haenszel.

Menurut Stark dan Chernyshenko (2002: 4), ada beberapa metode yang dapat digunakan untuk mendeteksi DIF, baik dengan statistika parametrik maupun statistika nonparametrik. Jika menggunakan metode statistika parametrik, maka dapat digunakan metode Khi-kuadrat dari Lord (*Lord's Chi-Square*), uji Perbandingan Kemungkinan (*Likelihood Ratio test*) dan metode Luas Daerah Bertanda dan Tidak Bertanda (*Signed and Unsigned Area Methods*), sedangkan untuk metode statistika nonparametrik, dapat digunakan metode Sibtest dan Mantel-Haenszel.

Selain mengklasifikasikan cara mendeteksi bias dengan statistik parametrik dan non parametrik, dapat pula dilihat berdasarkan pendekatan teori yang digunakan (Badrin Kartowagiran, 2004). Dengan menggunakan teori tes klasik, cara yang dapat ditempuh untuk mendeteksi bias butir adalah dengan analisis butir soal secara kuantitatif (membandingkan tingkat kesukarannya), dengan metode korelasi kelompok tunggal dan kelompok diferensial, metode diskriminasi butir, metode  $\chi^2$  dari Sceuneman dan Chamilli, metode Mantel-Haenszel, metode standardisasi, metode analisis distraktor, dan metode aturan emas (Chamilli & Shepart, 1994). Sedangkan berdasarkan teori respons butir, metode yang dapat digunakan untuk mendeteksi bias butir antara lain (1) kurve karakteristik butir (*item*

*characteristics curves* = ICC), (2) Khi-kuadrat menurut Lord, (3) metode luasan menurut Raju, dan (4) tes rasio kebolehjadian (*likelihood ratio test*).

Saat ini ada beberapa metode deteksi bias yang sedang dalam taraf pengembangan, di antaranya adalah : (a) metode liserel-komputer (liskom), (b) analisis model campuran (*mixture model analysis*), (c) tes bias secara simultan (*Simultaneous Item Bias Test* = SIBTEST), dan (d) model regresi logistik hirarki (*hierarchical logistic regression*).

## **KESIMPULAN**

Adanya bias butir pada suatu perangkat tes besar implikasinya dalam kehidupan siswa, mulai dari pemilihan sekolah pada jenjang studi lanjut, sampai memperoleh pekerjaan, mengingat peran penting matematika dan ilmu pengetahuan alam dalam kehidupan. Terkait dengan hal tersebut, perlu dideteksi muatan bias butir dalam suatu perangkat tes. Muatan bias ini dapat dideteksi dengan teori tes klasik maupun dengan teori respons butir.

## **DAFTAR PUSTAKA**

- Badrun, KW. 2004. Perbandingan metode pendeteksian bias menurut teori respons butir. *Disertasi*. Tidak dipublikasikan.
- Barrett, S. 2001. Differential item functioning: A case study from first year economics. *International Educational Journal Vol. 2, No. 3*. Diambil dari: <http://www.flinders.education/iej> tanggal 5 April 2003.
- Camilli, G. & Shepard, L.A. 1994. *Methods for identifying biased test items, Vol.4*. London: Sage Publications, Inc.
- Depdiknas. 2002. *Kurikulum 2004 standar kompetensi untuk mata pelajaran matematika*. Jakarta: Depdiknas.
- Gierl, M.J., Kholiq, S. & Boughton, K.1999. Gender Differential Item Functioning in Mathematics and Science: Prevalence and Policy Implications. *Paper*. presented at the Symposium entitled: "Improving Large-Scale Assessment in Education" at the Annual Meeting of the Canadian Society for the Study of Education. Quebec: Centre for Research in Applied Measurement and Evaluation University of Alberta. Sumber: <http://www.ucalgary.ca/iejll/volume7/burger.html> tanggal 12 Maret 2003.
- Grondlund, N.E. 1982. *Constructing Achievement Test* (3rd. ed. ). Englewood Cliffs, NJ: Prentice Hall Inc.
- Hambleton, R.K. & Rogers, H.J. 1995. Developing an item bias review form. Diambil dari: <http://www.ericcae.net/ft/tamu/biaspub2.htm> tanggal 10 Maret 2003.

- Jones, K., Evans, C., Byrd, R., & Campbell, K. 2000. Gender equity training and teacher behavior. *Journal of instructional psychology*, Sept 2000. Diambil dari: <http://www.findarticles.com/p/articles/mi> tanggal 16 Maret 2004.
- Kartini, K. 1980. *Psikologi Umum*. Jakarta: Kosgoro.
- Keeves, J.P. 1992. *The IEA technical handbook*. The Hague : The International Association for the Evaluation of Educational Achievement (IEA).
- Lee, V.N. 1999. Identifying differential item functioning on the NELS:88 History achievement test. *CSE Technical Report 511*. Diambil dari: <http://www.cse.ucla.edu/CRESST/Report/Tech5112pdf> tanggal 21 Mei 2003.
- Mackey, R.A., Diemer, M.A., & O'Brien, B.A. 2000. Psychological intimacy in the lasting relationships of heterosexual and same-gender couples. *Sex roles: A Journal of research*, August 2000. Diambil dari: <http://www.findarticles.com/p/articles/mi> tanggal 16 Maret 2004.
- Nowell, A., & Hedges, L.V. 1998. Trends in gender differences in academic achievement from 1960 to 1994: an analysis of differences in mean, variance, and extreme scores. *Sex roles: A Journal of research* July 1998. Diambil dari <http://www.snr-jnt.org/journalNT/JNT.html> tanggal 21 Juli 2004.
- Nurcholis. 1999. Fungsi butir diferensial pada perangkat tes Ebtanas Matematika SMU di D.I. Yogyakarta. *Tesis master*, tidak diterbitkan, Program Pascasarjana, Universitas Negeri Yogyakarta.
- Osterlind, S.J. 1983. *Test item bias*. Beverly Hills, CA: Sage Publications, Inc.
- Polk, D.L. & Erdal, K. 2002. Gender and sport differences in the emphasis on Physical and mental aspects of athletics. *Journal of Psychological Inquiry* vol 7, 2002. Diambil dari: <http://puffin.creighton.edu/psy/journal/vol7.html> tanggal 22 Juli 2004.
- Powel J.L. 2002. Variance in childrens perceptions of aggression as function of gender. *Journal of Psychological Inquiry* vol 7, 2002. Diambil dari: <http://puffin.creighton.edu/psy/journal/vol7.html> tanggal 22 Juli 2004.
- Scheuneman, J.D. & Blestein, C.A. 1999. Item Bias. *Advances in measurement in educational research and assessment*. (edit. Masters, G.N. dan Keeves, J.P.). Singapore: Pergamon, Pp. 220 – 234.
- Stark, S. & Chernishenko, O. 2002. *Detection of differential item/test functioning (DIF/DTF)*. University of Illinois at Urbana Champaign online. Diambil dari: <http://www.work.psych.ujuc.edu/irt/> tanggal 10 Februari 2003.