

## DETECTION OF GENDER BIASED USING DIF (DIFFERENTIAL ITEM FUNCTIONING) ANALYSIS ON ITEM TEST OF SCHOOL EXAMINATION YOGYAKARTA

Ajeng Wahyuni  
IAIN Ponorogo

### Alamat Korespondensi

[ajeng@iainponorogo.ac.id](mailto:ajeng@iainponorogo.ac.id)

s

### ABSTRACT

This study aim to detect an item that biased on gender (divided by male and female) by DIF (Differential Item Functioning) on school examination test in Yogyakarta regency in islamic education subject for third grade of Senior High School based on item respons theory . This research is quantitative research and the subject for this study are 937 Senior High School students in Yogyakarta regency taken by their response sheets on 50 items test with of choices of islamic education subject in school examination at year 2016. Data analysis based on Rasch model with Winstep 3.73 software. The research found that the there are item that biased on certain gender, there are 26 items that have probabilitas value less than 0.05 and t-value outside the range of  $-1.96 \leq t\text{-value} \leq 1.96$ , based on plot from the result analysis there are 14 items (6, 9, 10, 13, 15, 16, 20, 25, 28, 30, 31, 32, 41, 48) tend to give advantage for Male, and there are 12 items (1, 11, 12, 14, 21, 22, 29, 38, 40, 43, 46, 50) tend to give advantage for Female.

**Keywords:** DIF, Gender, School Examination, Rasch, Items Analysis

### 1. Pendahuluan

'Apakah pembelajaran tersebut berhasil?'. Pertanyaan tersebut muncul setelah proses pembelajaran. Untuk menjawab pertanyaan tersebut diperlukan sebuah evaluasi dan penilaian dalam suatu pembelajaran. Ebel dkk. Menyatakan tujuan dari evaluasi adalah untuk membuat judgement atau keputusan mengenai kualitas dari suatu hal dalam hal ini proses pembelajaran, dan pencapaian peserta didik (Ebel & Frisbie, 1991).

Penilaian merupakan suatu komponen dalam pembelajaran. Selain untuk mengetahui keberhasilan suatu proses pembelajaran, assessment juga berfungsi untuk mengetahui sampai dimana pemahaman peserta didik, membantu mengidentifikasi masalah dalam pembelajaran, dan membantu seorang guru memutuskan apakah perlu mengubah arah dan teknik pembelajaran (Barkley & Major, 2016a). Dan tentu saja, syarat yang terpenting dalam sebuah assessment adalah akuntabilitas. Salah satu alat assesemen yang kemudian sering digunakan dalam penilaian adalah tes baik tes pilihan ganda maupun esai.

'apakah tes tersebut valid dan reliabel? Apakah tes tersebut bias?' beberapa pertanyaan tersebut muncul setelah sebuah instrumen tes dibuat. Tentu saja, untuk mendapatkan hasil assessment yang akuntabel

tentu sebuah tes harus valid, reliabel dan terhindar dari bias. Kementerian Pendidikan juga memberikan aturan dalam Standar Penilaian Pendidikan Permendikbud No. 23 Th. 2016 bahwa assesmen haruslah bersifat adil dan objektif.

Apa yang dimaksud sebuah tes harus terhindar dari bias? Artinya tes tersebut tidak boleh memberikan suatu kelompok peserta tes tertentu keuntungan entah itu berdasarkan latar belakang peserta, suku, jenis kelamin dan lain-lain. Jika suatu kelompok tertentu diuntungkan ketika menjawab tes, maka tes tersebut bias dan harus diperbaiki. Dan sebuah tes yang baik tentu tidak boleh menguntungkan satu kelompok tertentu.

Untuk menjawab apakah sebuah tes tersebut bias atau tidak, dibuktikan dengan Differential Item Functioning (DIF). 'DIF Referes to items having different relative difficulty for groups and therefore violating invariance' (Andrich & Marais, 2019b), DIF bukan berarti menunjukkan bahwa suatu kelompok memiliki skor yang lebih besar dari kelompok lainnya, namun menunjukkan suatu item memiliki tingkat kesulitan yang berbeda pada grup yang berbeda sehingga melanggar asumsi invariansi parameter.

Salah satu cara untuk melihat DIF atau bias adalah dengan menggunakan Rasch Model yang

diperkenalkan pada tahun 1980 oleh George Rasch. 'The Rasch model is a theoretical framework for the analysis of measurement in studies of learning, intelligence, or ability' (Joughin, 2009) artinya bahwa Model Rasch adalah suatu kerangka teoritis untuk menganalisis pengukuran dalam sebuah pembelajaran. Selain itu Andersen menjelaskan Rasch Model sebagai suatu model analisis item dengan kurva karakteristik item logistic dengan slope yang sama atau dengan daya beda yang konstan. Model ini menjelaskan bagaimana suatu nilai bergantung pada tingkat kemampuan individu dan kesulitan pribadi mereka dalam melaksanakan setiap tugas. Artinya kinerja suatu kelompok tertentu tidak sama dengan kelompok lain.

Untuk memahami bagaimana DIF bekerja, kita perlu mengalihkan perhatian kita dari siswa ke item. 'Item adalah pertanyaan sebenarnya yang ditanyakan dalam tes' (Barkley & Major, 2016b). Beberapa pertanyaan bersifat sulit sementara pertanyaan lainnya tidak memerlukan pemahaman atau pemikiran untuk menjawabnya.

'Misalnya, pertanyaannya; "Apa ibu kota Sulawesi Tenggara?" pertanyaan ini tentu akan mudah bagi siswa yang tinggal di daerah Sulawesi Tenggara, namun pertanyaan ini bisa menjadi sangat sulit untuk siswa di luar daerah tersebut. Hal ini disebabkan kedua kelompok ini memiliki tingkat kemampuan berbeda dalam mengidentifikasi ibu kota dan hal itu tergantung pada kemampuan pribadi mereka (Andrich & Marais, 2019a).

DIF ini kemudian tidak hanya digunakan untuk mengidentifikasi bias gender namun dapat digunakan untuk melihat keberpihakan item soal berdasarkan kelompok-kelompok tertentu misal berdasarkan nasionalitas, warna kulit, daerah, dan ras.

Banyak penelitian yang sudah menggunakan DIF untuk menganalisis berbagai instrument soal dalam berbagai bidang, mulai dari Pendidikan, Psikologi, dan bidang lainnya.

Ujian Akhir Sekolah di Yogyakarta merupakan sebuah tes yang kemudian akan menentukan bagaimana keberhasilan siswa selama 3 tahun menempuh pendidikan di tingkat SMA. Tentunya, tes harus benar-benar objektif dan terhindar dari bias selain harus memenuhi kriteria valid dan reliabel. Tes harus benar-benar bisa mengukur kemampuan siswa dengan akurat, tanpa memihak salah satu kelompok tertentu. Apakah tes ini sudah valid dan reliabel? Apakah tes ini sudah objektif? Tentu hal tersebut menjadi pertanyaan besar yang perlu diselidiki.

Oleh karena itu, pada penelitian ini pengujian DIF berdasarkan Rasch Model akan digunakan untuk memeriksa ada tidaknya bias gender dalam soal Ujian Akhir Sekolah untuk mata pelajaran PAI. Ada tidaknya DIF dianalisis dengan melihat nilai prob. untuk setiap item. Jika nilai prob. suatu item lebih kecil dari 0.05 maka item tersebut mengandung DIF atau terindikasi bias.

## 2. Metode Penelitian

Penelitian ini bertujuan untuk menentukan apakah soal Ujian Akhir Semester mata pelajaran Pendidikan Agama Islam (PAI) yang dilaksanakan oleh kelas XII pada tahun 2016 terindikasi bias. Karena tujuannya bukan menguji keterkaitan antar variabel melainkan mendeskripsikan populasi apa adanya, maka penelitian ini termasuk penelitian survey (Creswell, 2015).

Penelitian ini merupakan penelitian populasi dan melibatkan 935 siswa yang mengikuti Ujian Akhir Semester untuk mata pelajaran PAI. Terdiri dari 420 siswa laki-laki dan 515 siswa perempuan.

**Tabel 1.** Populasi Penelitian

	Jumlah
<i>Laki-Laki</i>	420
<i>Perempuan</i>	515
	935

Soal Ujian Akhir Semester untuk mata pelajaran PAI pada tahun 2016 terdiri dari 50 soal yang dibuat oleh MGMP (Majelis Guru Mata Pelajaran) PAI Daerah Istimewa Yogyakarta. Data yang digunakan sebagai penelitian sudah mendapatkan izin dari Kementerian Agama Daerah Istimewa Yogyakarta.

Data mentah yang berasal dari Kementerian Agama Provinsi Yogyakarta ditransformasi ke dalam format excel dan notepad. Data kemudian dianalisis menggunakan Excel dan Winstep 3.73.

## 3. Hasil Penelitian dan Pembahasan

Analisis DIF menggunakan software Winstep memiliki dua output, output pertama berupa tabel identifikasi DIF dan grafik DIF. Pada tabel DIF, kita melihat nilai probabilitas masing-masing item jika suatu item memiliki nilai prob. lebih kecil dari 0.05 atau prob. < 0.05, maka dapat disimpulkan bahwa item tersebut mengandung DIF.

**Tabel 2 Hasil Analisis DIF**

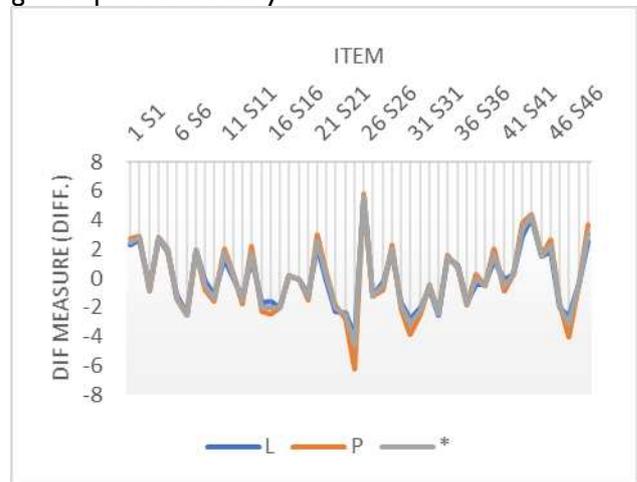
Item Soal	DIF Measure (L)	DIF Measure (P)	Prob.	Ket.
1	2.25	2.77	.0257	Bias
2	2.66	2.92	.3129	Tidak Bias
3	-.65	-.91	.0770	Tidak Bias
4	2.77	2.86	.7321	Tidak Bias
5	1.82	1.99	.3470	Tidak Bias
6	-1.11	-1.44	.0441	Bias
7	-2.48	-2.55	.7845	Tidak Bias
8	1.97	1.87	.5767	Tidak Bias
9	-.13	-.82	.0000	Bias
10	-1.15	-1.61	.0065	Bias
11	1.36	2.07	.0001	Bias
12	-.01	.26	.0435	Bias
13	-1.29	-1.74	.0113	Bias
14	1.43	2.16	.0001	Bias
15	-1.70	-2.28	.0052	Bias
16	-1.59	-2.43	.0001	Bias
17	-1.98	-2.03	.7782	Tidak Bias
18	.15	.15	.0003	Bias
19	-.10	-.07	.8705	Tidak Bias
20	-1.15	-1.51	.0132	Bias
21	2.17	2.99	.0008	Bias
22	-.18	.39	.0000	Bias
23	-2.26	-1.88	.0724	Tidak Bias
24	-2.36	-2.78	.1028	Tidak Bias
25	-4.03	-6.27	.0338	Bias
26	5.53	5.80	.7877	Tidak Bias
27	-1.21	-1.21	.0068	Bias
28	-.26	-.75	.0006	Bias
29	1.74	2.29	.0047	Bias
30	-1.63	-2.14	.0103	Bias
31	-2.78	-3.84	.0036	Bias
32	-2.00	-2.49	.0293	Bias
33	-.69	-.46	.104	Tidak Bias
34	-2.54	-2.38	.498	Tidak Bias
35	1.45	1.56	.4935	Tidak Bias
36	.87	.69	.2128	Tidak Bias
37	-1.66	-1.86	.2951	Tidak Bias
38	-.42	.27	.0000	Bias
39	-.43	-.51	.5648	Tidak Bias
40	1.24	2.02	.0000	Bias
41	-.09	-.91	.0000	Bias
42	.24	.24	.081	Tidak Bias
43	2.88	3.82	.0059	Bias
44	4.12	4.40	.5866	Tidak Bias
45	1.46	1.55	.604	Tidak Bias
46	1.89	2.60	.0008	Bias
47	-2.04	-1.77	.179	Tidak Bias
48	-2.67	-4.05	.0004	Bias
49	-.31	-.31	.0009	Bias
50	2.56	3.68	.0003	Bias

ringkasan item yang bias dan yang bebas dari bias.

**Tabel 3 Ringkasan Item DIF**

Description	Item	Total
Item DIF	1, 6, 9, 10, 11, 12, 13, 14, 15, 16, 20, 21, 22, 25, 28, 29, 30, 31, 32, 38, 40, 41, 43, 46, 48, 50	26
Item Not DIF	2, 3, 4, 5, 7, 8, 17, 18, 19, 23, 24, 26, 27, 33, 34, 35, 36, 37, 39, 42, 44, 45, 47, 49	24

Selanjutnya untuk melihat ke arah manakan item tersebut bias, dilakukan dengan melihat grafik DIF dan membandingkan kurva antara kurva yang melambangkan laki-laki, perempuan dan kurva normal. Pada Item yang baik atau objektif kurva ketiganya memiliki garis sejajar, sedangkan item yang terindikasi DIF akan memiliki kurva dimana satu kategori lebih tinggi dari kategori lainnya, kurva dengan posisi lebih tinggi menunjukkan bahwa item tersebut bias dan memberikan keuntungan terhadap kurva yang merepresentasikannya.



**Gambar 1: Kurva Analisis DIF**

Berdasarkan gambar 1 dapat kita perhatikan terdapat tiga garis dengan warna berbeda dimana garis biru merepresentasikan laki-laki, warna merah merepresentasikan perempuan dan warna abu-abu merepresentasikan kurva normal. Pada item yang bebas dari bias, ketiga titik warna tersebut bertumpang tindih pada satu titik yg sama. Namun, untuk item yg terindikasi bias terdapat titik yg lebih tinggi atau lebih rendah dari yang lainnya. Misalnya pada titik item 2, 3, dan 4 ketiga titik berada pada satu garis lurus dan ketiga titik saling bertumpang tindih, berbeda dengan item 1 dimana titik merah perempuan lebih tinggi dari titik normal dan titik biru lebih rendah dari titik normal. Pada kasus item nomor 1 karena titik merah yang melambangkan perempuan lebih tinggi maka dapat disimpulkan bahwa soal tersebut cenderung bias terhadap perempuan. Berikut ini hasil analisis pada item-item yang terindikasi bias.

Berdasarkan tabel diatas terlihat bahwa terdapat 26 item terindikasi DIF atau bias terhadap gender tertentu. Adapun tem yang memiliki nilai prob. lebih kecil dari 0.05 tersebut adalah item no. 1, 6, 9, 10, 11, 12, 13, 14, 15, 16, 20, 21, 22, 25, 28, 29, 30, 31, 32, 38, 40, 41, 43, 46, 50. Sedangkan item yang memiliki nilai prob lebih besar daro 0.05 atau aman dari bias berjumlah 24 item yaitu item 2, 3, 4, 5, 7, 8, 17, 18, 19, 23, 24, 26, 27, 33, 34, 35, 36, 37, 39, 42, 44, 45, 47, dan item 49. Berikut ini tabel

Tabel 4. Hasil Analisis Bias Item

Group	Item DIF	Total Item DIF
Male	6, 9, 10, 13, 15, 16, 20, 25, 28, 30, 31, 32, 41, 48	14
Female	1, 11, 12, 14, 21, 22, 29, 38, 40, 43, 46, 50	12

Berdasarkan hasil analisis pada tabel 3, terdapat 14 item yang bias terhadap laki-laki atau memberikan keuntungan tertentu kepada laki-laki dalam menjawab soal. Sedangkan 12 item lainnya bias terhadap perempuan.

#### 4. Kesimpulan

Sebuah instrument penilaian hasil belajar seyogyanya bebas dari soal-soal yang bias, namun berdasarkan hasil temuan dari 26 soal dari 50 soal terindikasi DIF atau bias terhadap kelompok gender tertentu, 14 soal bias terhadap gender laki-laki dan 12 soal lainnya bias terhadap gender perempuan. Padahal soal ini merupakan soal Ujian Akhir Tentu saja hal ini seharusnya hal ini dihindari dalam pembuatan sebuah instrument penilain agar mendapatkan hasil yang diinginkan.

#### 5. Daftar Pustaka

- Andrich, D., & Marais, I. (2019a). *A Course in Rasch Measurement Theory (Measuring in the Educational, Social and Health Sciences)*. In Springer. Springer Nature Singapore Ltd. <https://doi.org/10.1016/b978-008043348-6/50005-x>
- Andrich, D., & Marais, I. (2019b). *A Course in Rasch Measurement Theory Measuring in the Educational, Social and Health Sciences*. Springer Nature Singapore Pte Ltd. <http://www.springer.com/series/13812>
- Barkley, E. F., & Major, C. H. (2016a). *Learning Assessment Techniques: A Handbook for College Faculty*. 480. <https://www.wiley.com/en-us/Learning+Assessment+Techniques%3A+A+Handbook+for+College+Faculty-p-9781119050896>
- Barkley, E. F., & Major, C. H. (2016b). *Learning Assessment Techniques*. Josey Bass.
- Creswell, J. (2015). *Riset Pendidikan: Perencanaan, Pelaksanaan ds Evaluasi Riset Kualitatif dan Kuantitatif*. Pustaka Pelajar.

Ebel, R. L., & Frisbie, D. A. (1991). *ESSENTIALS OF EDUCATIONAL MEASUREMENT* (Fifth Edit). Prentice-Hall.

Joughin, G. (2009). *Assessment, learning and judgement in higher education*. In *Assessment, Learning and Judgement in Higher Education*. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-8905-3>