

## Identifikasi Muatan *Differential Item Functioning* Pada Data Ujian Nasional Matematika

Samritin

Program Studi Pendidikan Guru Sekolah Dasar, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Buton  
Jl. Betoambari No. 36 Baubau, Indonesia  
samritin75@gmail.com

### Abstract

One of the test criteria used in ability testing is a test that is free from Differential Item Functioning (DIF). Tests containing DIF can harm certain groups and benefit certain groups as well. However, sometimes the standard tests used are unknowingly containing DIF. The National Examination must meet the DIF-free criteria to avoid erroneous decision-making regarding the results of the exam. The purpose of this study is to detect Differential Item Functioning (DIF) through a unidimensional-based item response theory approach using explorative quantitative research. The results of the national mathematics exam as data for detecting DIF. The analysis technique uses the Mantel-Haenszal and Logistic Regression. In addition, DIF detection focused on gender, namely groups of men and women. Estimation of item parameters was carried out with the help of BILOG-MG, and the identification of DIF content using the R program tool. The results of the analysis showed that of the 21 items analyzed using Mantel-Haenszal and Logistic Regression, one item (Number 27) was suspected to contain DIF. This item is likely or suspected to benefit the male group, because in completing it the students first illustrate it in the form of an image, making it easier to calculate the height of the building, so that a long enough analysis is needed to be able to answer it.

**Keywords:** item response theory; unidimension; mantel-haenszal; logistc regression, DIF

### Abstrak

Salah satu kriteria tes yang digunakan dalam pengujian kemampuan adalah tes yang bebas *Differential Item Functioning* (DIF). Tes yang mengandung DIF dapat merugikan kelompok tertentu dan menguntungkan kelompok tertentu pula. Namun, kadang-kadang tes standar yang digunakan tidak disadari menandung DIF. Ujian Nasional haruslah memenuhi kriteria bebas DIF guna menghindari kesalahan pengambilan keputusan tentang capaian hasil ujian. Tujuan dari penelitian ini yaitu untuk mendeteksi *Differential Item Functioning* (DIF) melalui pendekatan *item response theory* berbasis unidimensi dengan menggunakan penelitian kuantitatif jenis eksploratif. Hasil ujian nasional matapelajaran matematika sebagai data untuk mendeteksi DIF. Teknik analisis menggunakan *Mantel-Haenszal* dan Regresi Logistik. Selain itu, diteksi DIF difokuskan pada gender yaitu kelompok pria dan wanita. Estimasi parameter butir dilakukan dengan bantuan *BILOG-MG*, dan identifikasi muatan DIF menggunakan alat bantu program R. Hasil analisis menunjukkan bahwa dari 21 butir soal yang dianalisis dengan menggunakan Mantel-Haenszal dan Regresi Logistik diperoleh sebanyak satu butir (Nomor 27) diduga memuat DIF. Butir tersebut kemungkinan atau diduga menguntungkan kelompok laki-laki, karena didalam menyelesaikannya siswa terlebih dahulu mengilustrasikan dalam bentuk gambar, sehingga memudahkan dalam menghitung tinggi gedung, sehingga diperlukan analisis yang cukup panjang untuk dapat menjawabnya

**Kata kunci:** teori respon butir, unidemensi, mantel-haenszal, regresi logistik, DIF

Copyright (c) 2022 Samritin

---

Corresponding author: Samritin

Email Address: samritin75@gmail.com (Jl. Betoambari No. 36 Baubau, Indonesia)

Received 15 Agustus 2022, Accepted 30 Agustus 2022, Published 30 Agustus 2022

## PENDAHULUAN

Pelaksanaan Ujian Nasional idealnya dilakukan secara obyektif, berkeadilan, dan akuntabel. Ujian nasional menggunakan instrument tes yang berkualitas. Butir soal tes yang digunakan tidak menguntungkan kelompok tertentu atau tidak memuat bias Retnawati (2013). Bias butir soal tes biasanya disebut sebagai *Differential Item Functioning* (DIF). DIF yaitu suatu item di dalam tes yang memiliki keberfungsian yang berbeda (Anggoff, 1982), sedangkan Hamblton et al., (1991)

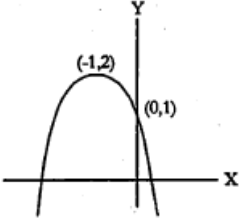
mengatakan bahwa ada kurva karakteristik butir tidak berhimpit pada sub kelompok berbeda. Menurut Penfield & Lam, 2000) bahwa ada dua jenis DIF yaitu DIF konsisten dan DIF tidak konsisten.

DIF pada butir soal dapat menguntungkan salah satu kelompok. Menurut Budiono (2005) menyatakan penyebabnya adanya DIF yaitu perbedaan fasilitas, kemampuan guru, dan tidak adilnya pelaksanaan tes. DIF bersifat diskriminatif (Retnawati, 2003). Mendeteksi DIF dapat dilakukan dengan cara membagi populasi menjadi dua kelompok (referensi dan pembanding). Gender, wilayah, dan etnis dapat dijadikan dasar dalam pembagian kelompok. Gender (Jenis kelamin) sering digunakan dalam mendeteksi DIF (Halpen, 1997). Penelitian tentang DIF sudah banyak dilakukan antara lain Curley & Schmitt (1993), Harris & Carlton (1993), Scheunerman & Gerritz (1990), Park (2008), Gallagher et al., (2002), Else-Quest et al., (2010), dan Nankervis (2011).

Kartowagiran (2005) mendeteksi keberadaan DIF pada UAN matematika SMP tahun 2003 di Yogyakarta dengan pengelompokan berdasarkan jenis kelamin. Retnawati (2003) mendeteksi keberadaan DIF pada tes masuk ditiga SMP di Yogyakarta tahun 2002 untuk mata ujian matematika dengan pengelompokan berdasarkan jenis kelamin. Penelitian yang dilakukan Budiono (2005) mendeteksi DIF pada UAN matematika tahun ajaran 2003/2004 di Kota Surakarta dengan pengelompokan berdasarkan jenis kelamin. Hasil penelitiannya menyimpulkan terdapat butir soal yang terkena DIF, Di bawah ini sajikan contoh butir soal tes yang terkena DIF menurut hasil penelitian Budiono (2005);

**Butir Soal Nomor 26:**  
**Persamaan grafik parabola pada gambar adalah ....**

a.  $y^2 - 4y + x + 5 = 0$   
 b.  $y^2 - 4y + x + 3 = 0$   
 c.  $x^2 + 2x + y + 1 = 0$   
 d.  $x^2 + 2x - y + 1 = 0$   
 e.  $x^2 + 2x + y - 1 = 0$



**Kunci jawaban: e**

Gambar 1. Butir Soal yang Diduga Memuat DIF

Karena memuat grafik, butir soal nomor 26 mengandung unsur spasial. Untuk dapat mengerjakannya, siswa terlebih dahulu harus mengetahui bentuk umum fungsi kuadrat yang mempunyai maksimum 2 di  $x = -1$ , kemudian melakukan substitusi untuk titik  $(0,1)$ . Jadi diperlukan analisis yang cukup panjang untuk dapat menjawabnya. Ada kemungkinan sebagian besar siswa perempuan tidak dapat mengerjakan butir soal ini karena tidak dapat melakukan analisis yang diperlukan. Jadi, terdeteksi butir soal 26 sebagai butir DIF yang menguntungkan kelompok laki-laki karena butir soal tersebut memuat unsur spasial (Budiono; 2005).

Mendeteksi DIF dapat menggunakan *item response theory* berunidimensi. Tingkat kesukaran, daya beda butir, dan tebakan semu disebut sebagai tiga parameter. Persamaan matematisnya sebagai berikut;

$$p_i(\theta) = c_i + 1 - (c_i) + \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$$

Keterangan:

$\theta$  = kemampuan siswa yang mengikuti tes

a = pembeda butir

b = kesukaran butir

c = tebakan semu

e = 2,718

Pi = Peluang siswa yang menjawab butir dengan benar

D = 1,7

Terdapat berbagai macam cara dalam mendeteksi DIF yaitu metode ANOVA, teori respon butir, analisis faktor, Mantel-Haenszel, model log-linear, dan Chi-square (Keeves, 1992), statistika parametrik dan non paramaterik (Stark et al., 2004). Pendeteksian DIF dalam penelitian ini menggunakan metode Mantel-Haenszel dan Regresi Logistik. Berikut persamaan matematis Mantel-Haenszel;

$$\alpha_{MH} = \frac{\sum_m \frac{R_{rm} W_{fm}}{N_{tm}}}{\sum_m \frac{R_{fm} W_{rm}}{N_{tm}}}$$

Keterangan:

$R_{fm}$  = kelompok fokus menjawab benar

$W_{fm}$  = kelompok fokus menjawab salah

$R_{rm}$  = kelompok acuan menjawab benar

$W_{rm}$  = kelompok acuan menjawab benar

$N_{tm}$  = banyaknya seluruh peserta tes pada level kemampuan.

$\alpha_{MH} > 1$  artinya menguntungkan kelompok acuan dan  $\alpha_{MH} < 1$  menguntungkan kelompok fokus.

Untuk menguji tingkat signifikansi DIF dapat menggunakan uji Chi-kuadrat sebagai berikut.

$$MH \chi^2 = \frac{\left[ \left| \sum_m R_{rm} - \sum_m E(R_{rm}) \right| - 0,5 \right]^2}{\sum_m Var(R_{rm})}$$

dengan:

dan

$$E(R_{rm}) = \frac{N_{rm} R_{rm}}{N_{tm}} \quad \text{Var}(R_{rm}) = \frac{N_{rm} R_{rm} N_{fm} W_{tm}}{N_{tm}^2 (N_{tm} - 1)}$$

Pendeksian DIF menggunakan metode Regresi Logistik mengikuti formula matematis sebagai berikut;

$$z = \delta + \tau_1 G + \tau_2 X + \tau_3 (GX)$$

Jika signifikan dan  $> 0$  maka dapat dikatakan bahwa butir tes memuat atau mengandung DIF yang menguntungkan kelompok acuan. Sebaliknya jika signifikan dan  $< 0$  maka menguntungkan kelompok fokus. Regresi logistik digunakan statistik uji Chi-Kuadrat Wald, yang dirumuskan sebagai berikut (Penfield, 2000);

$$W = \frac{\tau_k^2}{[SE(\tau_k)]^2}$$

## METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis eksplorasi untuk mendeteksi DIF pada soal tes. Data yang digunakan merupakan data skunder yaitu jawaban siswa SMP/MTs terhadap soal tes Ujian Nasional Tahun 2015 pada mata pelajaran matematika. Untuk membuktikan bahwa tes yang digunakan merupakan unidimensionalitas maka dilakukan uji asumsi unidimensi dengan menggunakan analisis faktor dilakukan dengan bantuan SPSS versi 20. Selanjutnya untuk mengestimasi parameter menggunakan BILOG-MG dan pendeteksian DIF butir digunakan metode *Mantel-Haenszal* dan metode Regresi Logistik.

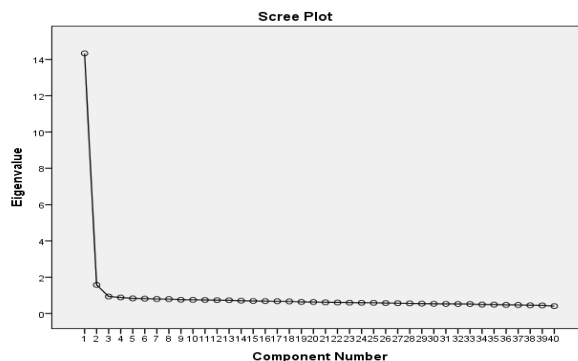
## HASIL DAN DISKUSI

Sebelum dilakukan pendetksian DIF, terlebih dahulu dilakukan pengujian asumsi unidimensi tes. Asumsi unidimensi tes didasarkan pada muatan faktor pada paket tes. Berdasarkan hasil analisis faktor menggunakan SPSS versi 20 yang disajikan pada Tabel 1, diperoleh hasil uji Bartlett sebesar 66050,043 yang menunjukkan nilai chi-kuadrat. Dengan derajat kebebasan 780 diikuti  $p\text{-value} < 0,01$ . Dengan demikian jumlah atau banyaknya sampel dapat dikatakan cukup.

Tabel 1. Hasil Uji KMO dan Bartlett

Kaiser-Meyer- Olkin Measure of Sampling Adequacy		0,988
Barlett's Test of Sphericity	Approx. Chi-Square	66050,043
	Df	780
	Sig.	0,000

Selanjutnya pembuktian keterpenuhan asumsi unidimensi tes dilakukan dengan analisis faktor. Analisis faktor dapat dibaca pada *scree-plot* dan nilai eigen pada varians yang terjelaskan. Hal ini dapat dilihat melalui *output* SPSS yang disajikan pada Gambar 2.



Gambar 2. Scree-Plot Nilai Eigen

Berdasarkan Gambar 2, dapat diduga bahwa perangkat tes yang diukur memuat dua faktor. Namun ada satu faktor yang memiliki nilai eigen sebesar 14,333 dengan persentase nilai varians yang terjelaskan sebesar 35,834%. Persentase ini sudah melebihi 20% sehingga perangkat tes yang diukur disimpulkan hanya memuat dimensi tunggal atau berifat unidimensi.

Tahapan selanjutnya memilih butir soal dengan pendekatan teori respon butir model logistik 3 paramater. Analisis parameter butir-butir tes menggunakan *BILOG-MG* sebagaimana telah disebutkan sebelumnya. Berdasarkan hasil analisis, diperoleh 21 butir dari 40 butir soal tes yang memenuhi model logistik 3 parameter. Butir-butir tes yang memenuhi model 3 parameter logistik tersebut beserta parameter butirnya disajikan pada Tabel 2.

Tabel 2. Parameter Butir dengan Pendekatan 3 Parameter Logistik

Butir	a	b	c
2	1.406	-0.685	0.207
6	1.956	-0.262	0.281
7	1.924	-0.204	0.283
10	1.753	0.508	0.326
12	0.863	0.500	0.305
13	1.051	-0.002	0.188
15	1.682	0.012	0.340
16	1.768	0.158	0.331
18	1.830	0.170	0.225
19	0.863	-0.154	0.192
23	0.872	0.290	0.234
24	1.379	0.305	0.221
27	1.703	-0.367	0.379
28	1.839	-0.041	0.308
29	1.946	0.355	0.272
30	1.285	-0.063	0.217
32	1.717	0.305	0.347
33	1.713	0.418	0.375
36	1.735	0.237	0.308
37	1.518	0.702	0.371
39	1.463	-0.036	0.387

Berdasarkan Tabel 2 di atas diketahui ada 15 butir yang tidak sesuai dengan model logistik 3 parameter, butir-butir tersebut yaitu butir nomor 1, 3, 4, 5, 8, 9, 11, 14, 17, 20, 21, 22, 25, 26, 31, 34, 35, 38, dan 40. Butir-butir tersebut dikeluarkan dari model. Selanjutnya butir-butir tes yang memenuhi model 3 parameter logistik tersebut digunakan untuk deteksi DIF. Untuk memudahkan pendeteksian DIF, estimasi parameter butir dikelompokkan ke dalam dua kelompok yaitu kelompok pria dan kelompok wanita.

Tabel 3. Parameter Berdasarkan Teori Respons Butir Unidimensi pada Kelompok Pria dan Wanita

Butir	Pria			Wanita		
	c	b	a	c	b	a
2	0.207	-0.608	3.307	0.207	0.803	1.447
6	0.281	-0.651	2.864	0.281	0.310	0.782
7	0.283	-1.215	1.713	0.283	0.924	1.725
10	0.326	-0.784	2.071	0.326	0.189	1.027
12	0.305	-0.285	1.515	0.305	0.414	0.644
13	0.188	-0.984	1.891	0.188	0.430	1.134
15	0.340	-0.506	1.742	0.340	1.137	0.677
16	0.331	-0.449	2.795	0.331	0.920	1.389
18	0.225	-0.672	0.943	0.225	0.646	0.624
19	0.192	-0.911	2.063	0.192	0.786	1.434
23	0.234	-0.120	2.048	0.234	1.086	0.996
24	0.221	-0.914	2.384	0.221	0.562	1.719
27	0.379	-1.009	1.407	0.379	0.714	1.171
28	0.308	-0.171	2.640	0.308	0.965	0.663
29	0.272	-0.431	1.676	0.272	0.367	0.774
30	0.217	-0.384	3.120	0.217	0.931	0.865
32	0.347	-0.172	1.780	0.347	1.402	0.943
33	0.375	-0.474	1.712	0.375	1.442	1.642
36	0.308	0.029	1.399	0.308	1.641	0.359
37	0.371	-1.555	1.385	0.371	0.880	1.157
39	0.387	-0.687	3.055	0.387	0.863	1.242

Pada Tabel 3 terlihat bahwa terdapat perbedaan hasil estimasi parameter setiap butir tes, baik parameter a maupun b. Sementara hasil estimasi parameter c tidak ditemukan ada perbedaan. Hasil estimasi parameter a, b, dan c tersebut digunakan sebagai acuan untuk mengidentifikasi DIF. Dari 21 butir tes yang dideteksi menggunakan metode *Mantel-Haenszal* dan Regresi Logistik dengan bantuan program R diperoleh hasil pada Tabel 4 dan Tabel 5.

Tabel 4. Hasil Deteksi DIF Menggunakan Metode *Mantel-Haenszal*

Butir	Stat.	P-value
b2	31.994	0.0737
b6	90.730	0.0026 **
b7	489.840	0.0048 **

b10	10.333	0.3094
b12	953.302	0.0020 **
b13	15.237	0.2171
b15	0.3967	0.5288
b16	0.4302	0.5119
b18	201.788	0.0085 **
b19	133.973	0.0023 **
b23	509.524	0.0078 **
b24	28.944	0.0889 .
<b>b27</b>	<b>1060.104</b>	<b>0.0000 **</b>
b28	304.689	0.0087 **
b29	137.615	0.0022 **
b30	110.400	0.0039 **
b32	25.074	0.0800 ***
b33	53.074	0.0012 *
b36	492.899	0.0092 **
b37	1.223.583	0.0010 **
b39	0.4983	0.4802

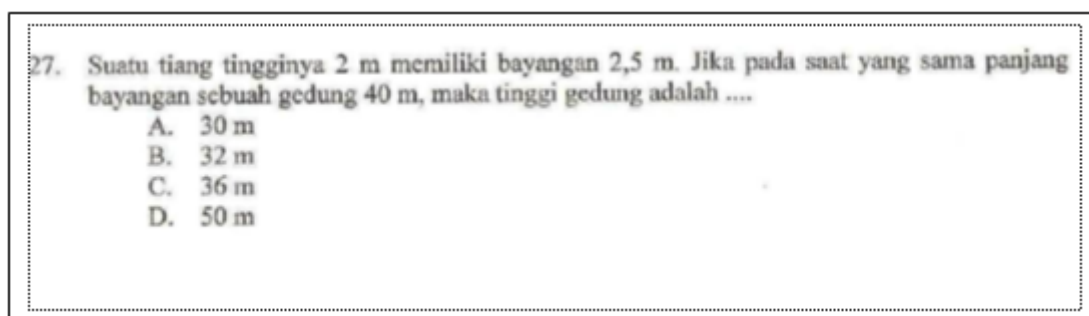
Berdasarkan hasil analisis menggunakan metode *Mantel-Haenszal* sebagaimana yang tersaji pada Tabel 4 di atas diduga terdapat 1 butir soal tes yang terdeteksi mengandung DIF. Butir tersebut secara statistik terindikasi signifikan memuat DIF karena memiliki  $p\text{-value} < 0$ . Butir soal dimaksud adalah butir nomor 27. Butir ini diduga menguntungkan kelompok acuan (laki-laki) karena memiliki  $\alpha\text{MH} > 1$ . Dengan demikian maka butir nomor 27 memiliki sifat diskriminatif atau tidak memenuhi kaidah adil dalam sistem pengujian. Selanjutnya deteksi DIF butir-butir tes juga dilakukan menggunakan metode Regresi Logistik. Hasil deteksi menggunakan metode Regresi Logistik dapat dilihat pada Tabel 5.

Tabel 5. Hasil Deteksi DIF Butir Menggunakan Metode Regresi Logistik

Butir	Stat.	P-value
b2	77.964	0.0023 **
b6	13.519	0.0051 **
b7	51.124	0.0037 **
b10	42.463	0.0032 **
b12	42.381	0.0034 **
b13	48.753	0.0040 **
b15	72.910	0.0022 **
b16	68.369	0.0036 **
b18	49.428	0.0845 .
b19	74.604	0.0014 **
b23	79.394	0.0015 **
b24	72.688	0.0017 **
<b>b27</b>	<b>111.373</b>	<b>0.0000 ***</b>
b28	26.317	0.0063 **
b29	49.428	0.0845 .
b30	48.000	0.0025 **
b32	46.371	0.0018 **
b33	97.859	0.0020 **
b36	44.670	0.0027 **
b37	88.577	0.0015 **

b39	47.497	0.0000 **
-----	--------	-----------

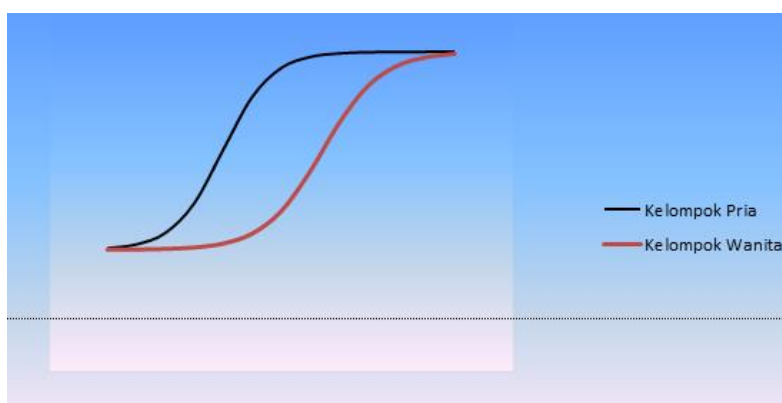
Berdasarkan hasil deteksi menggunakan metode *Mantel-Haenszal* sebagaimana yang tersaji dalam Tabel 5 diduga terdapat satu butir tes yang terdeteksi mengandung DIF. Butir tersebut secara statistik terindikasi signifikan memuat DIF karena memiliki  $p\text{-value} < 0$ . Butir ini diduga menguntungkan kelompok acuan (laki-laki) karena memiliki  $Z_t > 0$ . Hasil deteksi menggunakan metode Regresi Logistik sejalan dengan hasil hasil deteksi menggunakan metode *Mantel-Haenszal* yang mendeteksi butir yang sama sebagai butir mengandung DIF. Butir soal dimaksud adalah butir nomor 27. Dengan demikian maka butir nomor 27 memiliki sifat diskriminatif atau tidak memenuhi kaidah adil dalam sistem pengujian. Butir yang diduga mengandung DIF berdasarkan hasil deteksi metode *Mantel-Haenszal* dan metode Regresi Logistik disajikan pada Gambar 3.



Gambar 3. Butir Soal yang Diduga Memuat DIF

Butir soal nomor 27 di atas yang diduga menguntungkan kelompok laki-laki dapat disebabkan oleh proses penyelesaiannya yang perlu menghadirkan gambar ilustrasi. Sebelum menyelesaikannya, siswa terlebih dahulu mengilustrasikan dalam bentuk gambar, sehingga memudahkan dalam menghitung tinggi gedung. Karena butir soal nomor 27 mengandung unsur ilustrasi, maka diperlukan analisis yang cukup panjang untuk dapat menjawabnya. Ada kemungkinan sebagian besar siswa perempuan tidak dapat mengerjakan butir soal ini karena tidak dapat melakukan analisis yang diperlukan. Jadi, terdeteksi butir soal 27 sebagai butir DIF yang menguntungkan kelompok laki-laki. Temuan ini diperkuat dengan hasil penelitian (Park 2008).

Jika butir nomor 27 dibuat kurva karakteristik butirnya sebagai perbedaan peluang menjawab benar antara kelompok pria dan wanita maka diperoleh kurva sebagai berikut;



Gambar 4. DIF *uniform*



Pada Gambar 4, kurva karakteristik butir tidak saling memotong. Karena diketahui bahwa soal nomor 27 menguntungkan kelompok pria maka luas daerah peluang menjawab benar kelompok pria dikurangi kelompok wanita akan bernilai positif. Temuan penelitian ini sesuai dengan hasil penelitian yang dilakukan Budiono (2005).

## **KESIMPULAN**

Berdasarkan 40 butir soal diperoleh 21 butir yang digunakan untuk mendeteksi muatan DIF. Pemilihan butir tersebut menggunakan teori respon butir dengan pendekatan unidimensi. Berdasarkan uji kecocokan model diketahui bahwa model logistik 3 parameter yang memenuhi kriteria. Model tersebut memperhitungkan tingkat kesukaran butir, daya pembeda, dan tebakan semu. Hasil pengujian dengan menggunakan formula menggunakan *Mantel-Haenszel* dan *Regresi Logistik* diperoleh sebanyak 1 butir (Nomor 27) diduga memuat DIF. Butir tersebut kemungkinan menguntungkan kelompok laki-laki, karena didalam menyelesaikannya siswa terlebih dahulu mengilustrasikan dalam bentuk gambar, sehingga memudahkan dalam menghitung tinggi gedung. Ada kemungkinan sebagian besar siswa perempuan tidak dapat mengerjakan butir soal ini karena tidak dapat melakukan analisis yang diperlukan.

## **UCAPAN TERIMA KASIH**

Terimakasih diucapkan kepada Rektor Universitas Muhammadiyah Buton, Dekan Fakultas Keguruan dan Ilmu Pendidikan Universitas Muhammadiyah Buton, Ketua Program Studi Pendidikan Guru Sekolah Dasar Universitas Muhammadiyah Buton atas segala bantuan pembiayaan dan dukungan yang diberikan, serta seluruh tim yang membantu dalam pengumpulan dan analisis data.

## **REFERENSI**

- Angoff & William H. (1982). *Use of Difficulty and Discrimination Indices for Detecting Item Bias in Handbook of Methods for Detecting Test Bias*, ed. Ronard K. Beck. Baltimore: The Johns Hopkins University Press.
- Augemberg, K. E. & Morgan, D. L. (2008). *Differential Performance of Test Items by Geographical Regions*. Paper dipresentasikan pada annual meeting of the National Council on Measurement in Education, New York.
- Budiono. (2005). Perbandingan Metode Mantel-Haenszel, Sibtest, Regresi Logistik, dan Perbedaan Peluang Dalam Mendeteksi Keberbedaan Fungsi Butir. *Jurnal Studi Dan Evaluasi Pendidikan*, Vol. VII. No. 2. pp. 170-195.
- Curley, W. E., & Schmitt, A. P. (1993). *Revising SAT-Verbal Items to Eliminate Differential Item Functioning College Board Report*. No. 93-2, ETS RR No. 93-61. New York: The College Board

- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychological Bulletin*, 136(1), 103–127.
- Gallagher, A. M., Levin, J., & Cahalan, C. (2002). *Cognitive Patterns of Gender Differences on Mathematics Admissions Tests*. GRE Board Report No. 96- 17P. Princeton, NJ: Educational Testing Services.
- Hambleton, Ronald K., H. Swaminathan, dan H. Jane Rogers. (1991). *Fundamentals of Item Response Theory*. London: SAGE Publication, Inc.
- Halpern, Diana F. (1997). Sex Differences in Intelligence Implication for Education. *American Psychologist*, Vol. 52.
- Harris, A. M., & Carlton, S. T. (1993). Patterns of gender differences on mathematics items on the Scholastic Aptitude Test. *Applied measurement in Education*, 6(2), 137-151.
- Kartowagiran, B. (2005). *Perbandingan berbagai metode untuk mendeteksi bias butir* (Doctoral dissertation, Universitas Gadjah Mada).
- Keeves, J. (1992). *The IEA Study of Science III: Changes in science education and achievement: 1970 to 1984*. New York: Pergamon.
- Nandakumar, R. (1994). Assessing Dimensionality of a Set of Item Responses: Comparison of Different Approaches. *Journal of Educational Measurement*, 31(1), 17-35.
- Nankervis, B. (2011). Gender Inequities in University Admission due to the Differential Validity of the SAT. *Journal of College Admission*, 213, 24-30.
- Park, G.-P. (2008). Differential Item Functioning on an English Listening Test across Gender. *TESOL Quarterly*, 42(1), 115-123. doi: 10.2307/40264430
- Penfield, R. D. & Lam, T. C. M. (2000). Assessing differential item functioning in performance assessment: Review and recommendations. *Educational Measurement: Issues and Practice*, 19, 5-15.
- Retnawati, H. (2003). Keberfungsian Butir Diferensial pada Perangkat Tes Seleksi Masuk SLTP Mata Pelajaran Matematika. *Jurnal Penelitian dan Evaluasi Pendidikan*, 5(6).
- Retnawati, H. (2013). Pendeteksian keberfungsian butir pembeda dengan indeks volume sederhana berdasarkan teori respons butir multidimensi. *Jurnal penelitian dan evaluasi pendidikan*, 17(2), 275-286.
- Scheuneman, J. D., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, 27(2), 109-131.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the effects of differential item (functioning and differential) test functioning on selection decisions: When are statistically significant effects practically important?. *Journal of Applied Psychology*, 89(3), 497.