

# KOMPARASI METODE PENDETEKSIAN KETIDAKWAJARAN SKOR BERDASARKAN GENDER DAN RUANG LINGKUP MATERI PADA INSTRUMEN YANG MENGUKUR KEMAMPUAN PEMAHAMAN MATEMATIKA

Nur Hayati Setyaningsih  
SMK Negeri 3 Karawang

## ABSTRACT

*The purpose of this study is to compare the comparison of the inappropriateness index of scores by using the Jacob and Donlon Fisher method on high school math comprehension ability test on class XII science. The method adopted in this study is quantitative comparative. The subjects were participants' test scores, while the test participants are survey respondents. The population in this study are the scores of all test-takers on class XII in public high schools in Karawang. The samples in this study are 400 scores of class XII IPA students. Samples determination was performed using random sampling and based on gender. Data collection techniques utilized in this study is instruments of Mathematical Comprehension with the scope of Algebra and Geometry materials compiled by the researchers themselves. The data are analyzed by using Z test for different inappropriate test scores proportion of male and female learners on the inappropriateness index of the score under the Jacob and Donlon-Fisher method on an instrument that measures the ability of mathematical comprehension. The results showed that there are significant differences in the proportion of inappropriateness index. Donlon-Fisher method is more sensitive than the Jacob method in detecting the inappropriateness in scores so that the results obtained are more accurate.*

## Keywords

*Inappropriateness Index, Jacob method, Donlon-Fisher method, Gender, Algebra and Geometry*

## ABSTRAK

Tujuan dari penelitian ini adalah untuk membandingkan indeks ketidakwajaran skor dengan menggunakan metode Jacob dan Donlon Fisher pada tes kemampuan pemahaman Matematika SMA kelas XII IPA. Metode yang digunakan dalam penelitian ini adalah metode komparatif. Subjek penelitian adalah skor tes peserta. Populasi dalam penelitian ini adalah skor semua peserta tes siswa kelas XII di Sekolah Menengah Atas di Karawang. Sampel dalam penelitian ini adalah 400 skor siswa kelas XII IPA. Penentuan sampel dilakukan menggunakan sampel acak dan berdasarkan jenis kelamin. Teknik pengumpulan data yang dipergunakan dalam penelitian ini adalah instrumen Pemahaman Matematika dengan ruang lingkup materi Aljabar dan Geometri yang disusun oleh peneliti sendiri. Data dianalisis dengan menggunakan uji Z untuk menghitung proporsi ketidakwajaran skor peserta didik laki-laki dan perempuan menggunakan metode Jacob dan Donlon-Fisher, pada instrumen yang mengukur kemampuan pemahaman Matematika. Hasil penelitian menunjukkan bahwa ada perbedaan yang signifikan dalam proporsi indeks ketidakwajaran. Metode Donlon Fisher lebih sensitif daripada metode Jacob dalam mendeteksi ketidakwajaran skor sehingga hasil yang diperoleh lebih akurat.

## Kata Kunci

Indeks Ketidakwajaran, Metode Jacob, Metode Donlon-Fisher, Jenis Kelamin, Aljabar dan Geometri

Alamat Korespondensi  
SMK Negeri 3 Karawang

e-mail:  
nur\_ning@yahoo.com

## I. Pendahuluan

Instrumen tes dalam kegiatan pembelajaran memiliki peran yang sangat penting, yaitu: sebagai alat bantu untuk mengungkap kemampuan peserta didik baik kognitif maupun psikomotorik. Menyadari peranan tersebut maka instrumen tes harus memenuhi persyaratan tes

yang baik, sehingga hasil yang diperoleh dapat menggambarkan kemampuan yang sesungguhnya dari peserta didik. Adakalanya hasil tes yang berupa skor tidak menggambarkan kemampuan yang sesungguhnya dari peserta didik sehingga terjadi ketidakwajaran skor.

Ketidakwajaran skor dapat terjadi karena berbagai faktor, misalnya kecerobohan peserta didik saat menjawab atau tindakan curang, seperti mencontek serta kecemasan. Artinya ketidakwajaran skor adalah skor yang diperoleh peserta tes yang tidak sesuai dengan kemampuan sebenarnya dari peserta tes, dengan asumsi semua butir tes sudah sesuai dengan kriteria soal yang baik (Naga, 2013).

Di dalam penskoran uji tes, perlu dideteksi adanya ketimpangan skor yang berbentuk ketidakwajaran ini. Pada teori skor, baik klasik maupun modern, pendeteksian ini dilakukan melalui perbandingan di antara hasil jawaban pada butir mudah dengan hasil jawaban pada butir sukar.

Adalah wajar kalau peserta dapat menjawab dengan benar butir yang mudah dan tidak dapat menjawab dengan benar butir yang sukar. Dan adalah tidak wajar kalau sebaliknya peserta ujites itu berhasil menjawab dengan benar butir yang sukar dan tidak dapat menjawab dengan benar butir yang mudah. Indeks ketidakwajaran skor dapat dihitung melalui pendekatan teori pengukuran klasik, yaitu: dengan metode Ghiselli, metode Jacob, metode Donlon Fisher dan metode Sato yang dimodifikasi oleh Harnisch dan Linn yang selanjutnya disingkat metode SHL.

Pembelajaran Matematika rekomendasi *National Council of Teachers of Mathematics* atau NCTM menggariskan bahwa peserta didik harus mempelajari Matematika melalui pemahaman dan aktif membangun pengetahuan baru dari pengalaman dan pengetahuan yang telah dimiliki sebelumnya. Dengan pemahaman, peserta didik dapat lebih mengerti akan konsep materi pelajaran itu sendiri.

Perbedaan jenis kelamin bukan hanya berakibat pada perbedaan kemampuan dalam Matematika, tetapi cara memperoleh pengetahuan

Matematika juga terkait. Berdasarkan beberapa ahli dibidang psikologis, mengatakan bahwa perempuan pada umumnya lebih baik pada ingatan dan laki-laki lebih baik dalam berpikir logis. Jadi besar kemungkinan adanya perbedaan antara peserta didik laki-laki dan perempuan dalam kemampuan pemahaman Matematika untuk materi Aljabar dan Geometri.

Penelitian terdahulu yang relevan dengan penelitian ini diantaranya penelitian yang dilakukan oleh Suciati Rahayu Widyastuti yang menyimpulkan bahwa tidak terdapat perbedaan indeks ketidakwajaran skor dengan menggunakan metode SHL dan metode Donlon Fisher terhadap hasil tes belajar Matematika SMP kelas VII, kemudian penelitian selanjutnya oleh Mugia Rahayu Awwalunnisa yang menyimpulkan bahwa dengan metode Jacob indeks ketidakwajaran skor pada soal dengan lima pilihan jawaban lebih kecil secara signifikan dibanding dengan skor tidak wajar pada soal dengan tiga pilihan jawaban dan penelitian yang dilakukan oleh S. Fedi, dkk pada tahun 2014 menyatakan adanya perbedaan tingkat kecemasan Matematika antara peserta didik laki-laki dan perempuan, di mana perempuan lebih cemas daripada laki-laki sehingga tingkat apresiasi Matematika berbeda juga.

Dari uraian tersebut maka perlu adanya pengkajian komparasi indeks ketidakwajaran skor metode Jacob dan metode Donlon Fisher pada peserta didik berdasarkan gender dan ruang lingkup materi dengan menggunakan instrumen yang mengukur pemahaman Matematika di Sekolah Menengah Atas.

Tujuan dari penelitian ini adalah untuk mengetahui komparasi metode pendeteksian ketidakwajaran skor dengan metode Jacob dan metode Donlon Fisher berdasarkan gender dan ruang lingkup materi pada instrumen yang mengukur kemampuan pemahaman Matematika.

## 2. METODE PENELITIAN

Penelitian ini termasuk penelitian quasi eksperimen dengan menggunakan model perbandingan proporsi. Variabel dalam penelitian ini ada dua, yaitu: variabel bebas dan variabel

terikat. Variabel bebas dalam penelitian ini adalah *gender* dan ruang lingkup materi. Gender dalam penelitian ini adalah laki-laki dan perempuan sedangkan ruang lingkup materi dalam penelitian ini adalah Aljabar dan Geometri. Sedangkan variabel terikatnya adalah proporsi skor tidak wajar berdasarkan indeks ketidakwaajaran yang dihitung berdasarkan teori klasik dengan menggunakan metode Jacob dan metode Donlon Fisher.

Rancangan penelitian yang digunakan adalah penelitian komparatif, yaitu: membandingkan proporsi berdasarkan indeks ketidakwaajaran skor yang diperoleh. Model rancangan ini dipilih karena indeks ketidakwaajaran yang digunakan adalah indeks ketidakwaajaran baku sehingga yang dibandingkan adalah proporsi skor tidak wajar berdasarkan indeks ketidakwaajaran yang diperoleh.

Prosedur penelitian yang dilaksanakan meliputi tiga tahap, yaitu: tahap persiapan, tahap pelaksanaan, dan tahap analisis. Masing-masing tahap akan diuraikan sebagai berikut:

- a. Tahap persiapan, kegiatan yang dilakukan pada tahap ini meliputi: (1) menyusun instrumen kemampuan pemahaman Matematika kelas XII IPA SMA; (2) menentukan sekolah yang akan dijadikan sebagai tempat penelitian; (3) meminta izin pelaksanaan penelitian kepada kepala sekolah; (4) melakukan validasi instrumen yang mengukur kemampuan pemahaman Matematika yang telah dibuat kepada dosen pendidikan Matematika, mahasiswa penelitian dan evaluasi pendidikan serta guru Matematika SMA/SMK; (5) uji coba instrumen; (6) menghitung reliabilitas, tingkat kesukaran, daya beda, dan distraktor dengan bantuan *excel* dan *iteman*; (7) revisi instrumen; dan (8) membuat kesepakatan dengan guru bidang studi Matematika mengenai waktu dan kelas yang akan digunakan untuk penelitian.
- b. Tahap pelaksanaan, kegiatan yang dilakukan pada tahap ini meliputi: (1) memilih sampel penelitian dan (2) memberikan tes kemampuan pemahaman Matematika meliputi ruang lingkup materi Aljabar dan Geometri yang berupa tes pilihan ganda dengan lima alternatif jawaban.

- c. Tahap analisis data, kegiatan yang dilakukan dalam tahap ini adalah: (1) mengkoreksi dan menghitung skor dari 400 jawaban peserta didik; (2) menghitung indeks ketidakwaajaran skor responden dengan metode Jacob dan metode Donlon Fisher dengan menggunakan *excel*; (3) menghitung transformasi ke skor baku T dengan bantuan *excel*; (4) menghitung skor tidak wajar untuk masing-masing kelompok data yang diperlukan dan proporsinya; (5) menghitung statistika deskriptif indeks kewajaran dengan *SPSS 17 for Windows*; (6) melakukan uji prasyarat analisis yang meliputi uji normalitas dan uji homogenitas dengan *SPSS 17 for Windows*; dan (7) melakukan uji hipotesis dengan menggunakan uji Z.
- d. Pembuatan kesimpulan, pada tahap ini peneliti membuat kesimpulan berdasarkan data yang diperoleh.

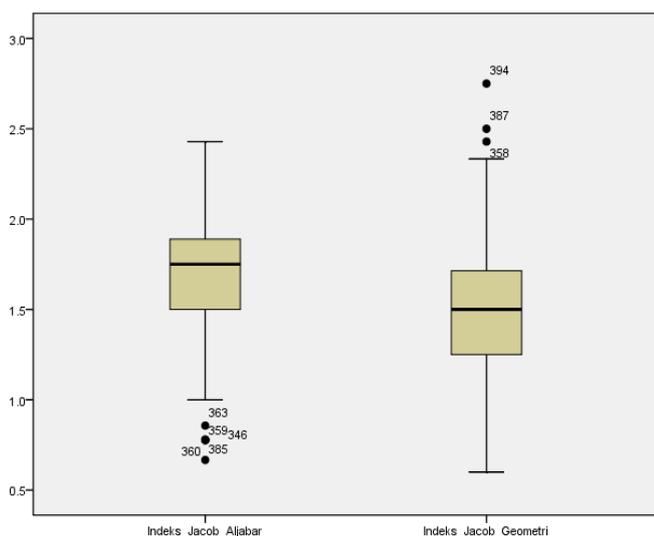
### 3. HASIL DAN PEMBAHASAN

Uji validitas isi dilakukan oleh lima pakar yang merupakan dosen Matematika, dosen pengukuran, dan guru Matematika SMA/SMK. Perhitungan kecocokan dihitung dari hasil validasi pakar/panelis yang terdiri atas dosen Matematika, dosen pengukuran, dan guru Matematika SMA/SMK yang berjumlah 23 orang dan dianalisis dengan menggunakan rumus rasio validitas isi atau *content validity ratio* dari Lawse yang sering disingkat menjadi CVR. Dari hasil analisis diperoleh bahwa dari 50 butir soal ujicoba baik karena nilai CVR lebih dari nol ( $CVR > 0$ ).

Uji coba instrumen kemampuan pemahaman Matematika yang terdiri atas 50 butir soal dilakukan kepada peserta didik kelas XII SMA jurusan IPA sebanyak 279 orang, dengan reliabilitas 0,893 dan ada 14 butir drop sehingga diperoleh 36 butir valid yang memiliki tingkat kesukaran dan daya beda yang proporsional serta distraktor yang berfungsi dengan baik.

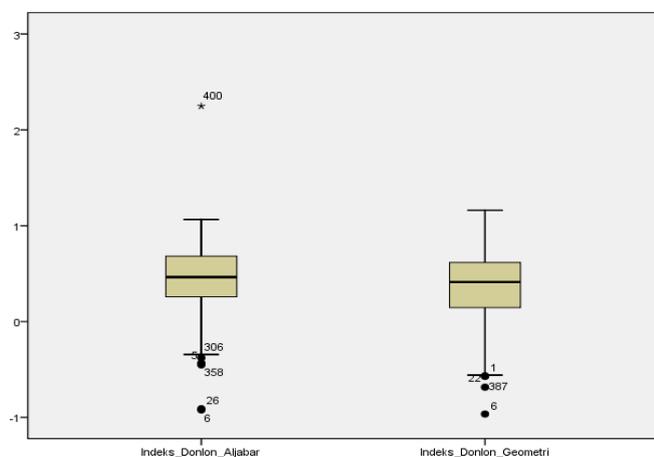
Penelitian ini menggunakan 400 sampel yang terdiri dari 200 peserta didik laki-laki dan 200 peserta didik perempuan. Dari hasil penelitian diperoleh skor pemahaman Matematika

materi Aljabar dan Geometri yang kemudian dihitung indeks ketidakwaajaran skornya dengan metode Jacob dan Donlon Fisher, baik pada peserta didik laki-laki maupun perempuan. Hasil indeks ketidakwaajaran skor dengan kedua metode disajikan dalam bentuk *boxplot* sebagai berikut:



Gambar 1. Boxplot Indeks Ketidakwaajaran Skor Metode Jacob

Dari Gambar 1 terlihat bahwa indeks ketidakwaajaran skor kemampuan pemahaman Matematika berdasarkan metode Jacob pada ruang lingkup materi Aljabar memiliki pencilan bawah sedangkan pada materi Geometri terdapat pencilan atas. Rata-rata indeks ketidakwaajaran skor terlihat lebih tinggi pada materi Aljabar dibandingkan materi Geometri.



Gambar 2. Boxplot Indeks Ketidakwaajaran Skor Metode Donlon Fisher

Dari Gambar 2 terlihat bahwa indeks ketidakwaajaran skor kemampuan pemahaman Matematika berdasar metode Donlon Fisher pada ruang lingkup materi Aljabar dan Geometri memiliki pencilan bawah. Rata-rata indeks ketidakwaajaran skor ruang lingkup materi Aljabar dan Geometri terlihat hampir sama.

Berdasarkan hasil kesimpulan uji hipotesis dengan uji Z maka pada hipotesis 1 sampai 4 ternyata dugaan peneliti tidak tepat dalam menyimpulkan dugaan sementara mengenai adanya perbedaan proporsi skor tidak wajar berdasarkan indeks ketidakwaajaran skor metode Jacob dan Donlon Fisher untuk instrumen pemahaman Matematika materi Aljabar dan Geometri pada peserta didik laki-laki dan perempuan. Pada keempat hipotesis tidak ada perbedaan proporsi yang signifikan antara peserta didik laki-laki dan perempuan.

Ketidakwaajaran skor pada tes Matematika dapat terjadi pada peserta tes yang pada saat pelaksanaan merasa cemas, was-was, belum siap menghadapi tes karena belum belajar, terburu-buru menjawab, salah menafsirkan soal karena jarang latihan soal, mencontek, atau menebak jawaban soal. Kondisi tersebut juga terjadi pada peserta tes baik laki-laki dan perempuan yang menjadi responden penelitian, karena tes dilaksanakan tanpa pemberitahuan terlebih dahulu. Dalam penelitian ini kondisi peserta didik baik laki-laki maupun perempuan dapat sama-sama cemas atau sama-sama kondisi siap karena pelaksanaan pengambilan data satu minggu menjelang Ujian Nasional. Hal tersebut sesuai dengan hasil penelitian Eti Nurhayati yang menyatakan bahwa terdapat pengaruh yang signifikan antara tingkat kecemasan dalam menghadapi ujian terhadap hasil belajar Matematika peserta didik dan juga hasil penelitian Nasmiawati menyatakan bahwa terdapat pengaruh positif kecemasan terhadap ketidakwaajaran skor.

Ketidakwaajaran skor yang terdeteksi tidak signifikan kemungkinan juga karena efek tebakan dari peserta didik laki-laki maupun perempuan, hal tersebut terjadi karena instrumen pemahaman Matematika berupa soal pilihan ganda dan dalam

pengambilan data disampaikan bahwa penskoran menggunakan *correct score* sehingga tidak ada pengurangan nilai jika jawaban salah.

Indeks ketidakwaajaran dengan menggunakan metode Jacob diperoleh dengan menggunakan rata-rata bobot. Butir soal dikelompokkan dalam lima kelompok berdasarkan tingkat kesukarannya. Dari kelima kelompok soal tersebut diberi bobot nol untuk kelompok butir termudah dan secara berurutan sehingga pada kelompok lima atau yang tersulit diberi bobot empat. Pengelompokan dalam perhitungan indeks Jacob sudah berdasarkan tingkat kesukaran dengan *judgement* peneliti dengan seproporsional mungkin untuk mengurangi kelemahan metode Jacob tentang pembagian butir dalam lima kelompok, sehingga hasil skor tidak wajar merupakan perhitungan yang mencerminkan keadaan yang sebenarnya.

Menurut analisis terbaru kemampuan Matematika peserta didik perempuan tidak lebih buruk daripada laki-laki. Pada materi Aljabar hasil penelitian Larbi (2014) menyatakan bahwa tidak ada perbedaan yang signifikan antara peserta didik laki-laki dan perempuan dalam menyelesaikan masalah Aljabar. Hal tersebut mendukung hasil uji hipotesis 1 dan 3 bahwa tidak terdapat perbedaan proporsi skor tidak wajar yang signifikan antara peserta didik laki-laki dan perempuan pada materi Aljabar. Kemudian hasil penelitian Mainali (2014) menunjukkan tidak terdapat perbedaan yang signifikan kemampuan Geometri dari aspek gender. Jadi diterimanya hipotesis 2 dan 4 selaras dengan hasil penelitian tersebut.

Hasil uji hipotesis 5 menunjukkan bahwa proporsi skor tidak wajar berdasar metode Donlon Fisher lebih rendah daripada metode Jacob untuk peserta didik perempuan pada materi Aljabar atau dengan kata lain tidak ada perbedaan proporsi ketidakwaajaran skor peserta didik perempuan berdasar metode Jacob dan metode Donlon Fisher pada ruang lingkup Aljabar. Tidak adanya perbedaan proporsi terjadi bukan karena masalah instrumen pemahaman Matematika yang digunakan dalam penelitian, sebab reliabilitas instrumen sudah 0,893 atau mendekati 0,9 sehingga reliabel atau ajeg untuk menilai

kemampuan peserta didik pada ruang lingkup materi Aljabar. Perhitungan indeks ketidakwaajaran skor metode Jacob pada skor materi Aljabar juga sudah diminimalkan kelemahannya dengan pengelompokan butir soal berdasarkan tingkat kesukaran dengan proporsional, sehingga optimal dalam pemberian bobot tiap kelompok butir soal.

Dari hasil uji hipotesis 6, 7, dan 8 menunjukkan bahwa proporsi skor tidak wajar berdasarkan indeks ketidakwaajaran skor metode Donlon Fisher lebih tinggi dari pada metode Jacob pada ruang lingkup materi Aljabar untuk peserta didik laki-laki dan materi Geometri untuk peserta didik laki-laki dan perempuan. Metode Donlon Fisher lebih cermat dalam mendeteksi ketidakwaajaran skor berdasarkan kesukaran butir dalam skala delta. Dalam perhitungan indeks ketidakwaajaran skor metode Donlon Fisher dengan lebih rumit dan membutuhkan waktu yang lebih lama karena memerlukan rumus yang bertingkat dibandingkan metode Jacob.

Pendeteksian ketidakwaajaran skor dengan data yang banyak dengan harapan data tersebut berdistribusi normal, hendaknya menggunakan metode Donlon Fisher karena lebih efektif dan akurat. Metode Donlon Fisher dan metode Jacob memiliki dasar perhitungan kesukaran butir soal. Pada metode Jacob lebih sederhana karena hanya dengan pengelompokan dan pembobotan namun pada perhitungan metode Donlon Fisher menyatakan kesukaran butir dalam skala Delta yang merupakan taraf kesukaran butir yang ditransformasikan ke dalam distribusi normal baku. Kesukaran butir soal yang dinyatakan dalam skala Delta memang menghasilkan bentuk kesukaran butir soal yang baru, namun tidak mengubah besaran kesukaran butir soal.

Dari hasil uji hipotesis 9 dan hipotesis 10 menunjukkan bahwa pada ruang lingkup materi Aljabar maupun Geometri ada perbedaan proporsi skor tidak wajar yang signifikan di mana metode Donlon Fisher lebih cermat daripada metode Jacob. Metode Donlon Fisher mendeteksi ketidakwaajaran skor menggunakan korelasi biserial untuk membandingkan pola kesukaran peserta dengan pola kesukaran kelompok artinya menggunakan rumus yang bertingkat dibandingkan

dengan metode Jacob yang perhitungannya didasarkan pada pengelompokan tingkat kesukaran dan pembobotan yang merupakan *judgement* peneliti.

Instrumen pemahaman Matematika yang terdiri dari ruang lingkup Aljabar dan Geometri mampu ajeg mengukur kemampuan peserta didik kelas XII SMA IPA baik laki-laki maupun perempuan sehingga skor yang dihasilkan dapat dideteksi ketidakwajaran skornya dengan metode Jacob dan Donlon Fisher. Indeks ketidakwajaran skor yang dihasilkan mendukung hipotesis yang diajukan sehingga komparasi yang diuji menunjukkan adanya perbedaan proporsi yang signifikan. Metode Donlon Fisher lebih akurat dibanding metode Jacob untuk mendeteksi ketidakwajaran skor.

#### 4. KESIMPULAN

Berdasarkan hasil analisis data, pengujian prasyarat, pengujian hipotesis, dan pembahasan maka disimpulkan bahwa:

- Proporsi skor tidak wajar peserta didik perempuan berdasarkan indeks ketidakwajaran skor metode Donlon Fisher pada instrumen yang mengukur kemampuan pemahaman Matematika ruang lingkup materi Geometri lebih tinggi daripada metode Jacob dengan perbedaan proporsi yang signifikan.
- Proporsi skor tidak wajar peserta didik laki-laki berdasarkan indeks ketidakwajaran skor metode Donlon Fisher pada instrumen yang mengukur kemampuan pemahaman Matematika ruang lingkup materi Aljabar lebih tinggi daripada metode Jacob dengan perbedaan proporsi yang signifikan.
- Proporsi skor tidak wajar peserta didik laki-laki berdasarkan indeks ketidakwajaran skor metode Donlon Fisher pada instrumen yang mengukur kemampuan pemahaman Matematika ruang lingkup materi Geometri lebih tinggi daripada metode Jacob dengan perbedaan proporsi yang signifikan.
- Proporsi skor tidak wajar peserta didik laki-laki dan perempuan berdasarkan indeks

ketidakwajaran skor metode Donlon Fisher pada instrumen yang mengukur kemampuan pemahaman Matematika ruang lingkup materi Aljabar lebih tinggi daripada metode Jacob dengan perbedaan proporsi yang signifikan.

- Proporsi skor tidak wajar peserta didik laki-laki dan perempuan berdasarkan indeks ketidakwajaran skor metode Donlon Fisher pada instrumen yang mengukur kemampuan pemahaman Matematika ruang lingkup materi Geometri lebih tinggi daripada metode Jacob dengan perbedaan proporsi yang signifikan.

Jadi kesimpulan secara umum, perbedaan jenis kelamin/ gender tidak memberikan pengaruh pada ketidakwajaran skor pemahaman Matematika dan metode Donlon Fisher lebih sensitif daripada metode Jacob dalam pendeteksian ketidakwajaran skor peserta didik laki-laki dan perempuan pada instrumen pemahaman Matematika ruang lingkup Aljabar dan Geometri di Sekolah Menengah Atas.

#### 5. DAFTAR PUSTAKA

- Abu, Mohd. Salleh dan Zaid Zainal Abidin. "Improving the Levels of Geometric Thinking of Secondary School Students Using Geometry Learning Video based on Van Hiele Theor," *International Journal of Evaluation and Research in Education (IJERE)*, Vol.2, No.1, March 2013.
- Agustiar, Wisnawati, dan Yuli Asmi, "Kecemasan Menghadapi Ujian Nasional dan Motivasi Belajar Pada Siswa Kelas XII SMA Negeri "X" Jakarta Selatan," *Jurnal Psikologi*, Volume 8 Nomor 1, Juni 2010, h. 13.
- Awwalunnisa, Mugia Rahayu. "Pendeteksian Ketidakwajaran Skor Berdasarkan Metode Jacob Ditinjau Dari Jumlah Pilihan Jawaban Pada Tes Pilihan Ganda." Tesis: Universitas Pendidikan Indonesia, 2014.
- Drasgow, Fritz, Michael V. Levine dan Mary E. McLaughlin. "Detecting Inappropriate Test Scores with Optimal and Practical

- Appropriateness Indies." *Applied Psychological Measurement*, Vol. 11, No. 1, March 1987.
- Drasgow, Fritz. "Choice of Test Model for Appropriateness Measurement," *Applied Psychological Measurement*, Vol. 6 No. 3, Summer 1982.
- Geary, David C, Scott J. Saults, Fan Liu, and Mary K. Hoard. "Sex Differences in Spatial Cognition, Computational Fluency, and Arithmetical Reasoning." *Journal of Experimental Child Psychology*, Vol. 77, 2000.
- Geit, E. N. dan M. King, "Different, Not Better: Gender Differences in Mathematics Learning and Achievement," *Journal of Instruction Psychology*, Vol 35 (1), 2006.
- Hamzah, Ali. *Evaluasi Pembelajaran Matematika*. Jakarta: PT Raja Grafindo Persada, 2014.
- Harahap, Mahyudin dan Wardhani, Nugaan Yulia. "Pendeteksian Ketidakwajaran (*inappropriateness*) Skor Hasil Ujian Nasional Mata Pelajaran Matematika SLTA Se-kota Medan Tahun Pelajaran 2006/2007." *Tesis*, Universitas Indonesia, 2006.
- Naga, Dali S. *Pengantar Teori Skor pada Pengukuran Pendidikan*. Jakarta: Gunadarma, 1992.
- Naga, Dali S. *Teori Skor pada Pengukuran Mental*. Jakarta: PT. Nagarani Citrayasa. 2013.
- NCTM. *Executive Summary Principle and Standards for School Mathematics*. Reston, VA: NCTM, 2000.
- Susetyo, Budi. *Prosedur Penyusunan & Analisis Tes untuk Penilaian Hasil Belajar Bidang Kognitif*. Bandung: PT Refika Aditama. 2015.
- Usodo, Budi. "Karakteristik Intuisi Siswa SMA dalam Memecahkan Masalah Matematika Ditinjau dari Kemampuan Matematika dan Perbedaan Gender," *AKSIOMA*, Volume 01, Nomor 01 Maret 2012 (diakses 15 Januari 2016).
- Valentine, Elizabeth F. *Gender Differences in Learning and Achievement in Mathematics, Science, and Technology and Strategies for Equity: A Literature Review*. Virginia: ERIC, 2000.
- Widyastuti, Suciati Rahayu. "Efektivitas Metode SHL dan Donlon Fisher untuk Pendeteksian Ketidakwajaran Skor pada Hasil Belajar Matematika." *Prosiding Seminar Implementasi Kurikulum 2013*. 2014.
- Zubaidah, Amir MZ, "Perspektif Gender Dalam Pembelajaran Matematika," *Marwah*, Vol. XII, No. 1 Juni Th. 2013.