

**PERBEZAAN PENCAPAIAN BERDASARKAN
JANTINA DALAM PEPERIKSAAN
PERCUBAAN MATEMATIK
UPSR**

ABD WAHAB BIN MUDA

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2006

**PERBEZAAN PENCAPAIAN BERDASARKAN JANTINA
DALAM PEPERIKSAAN PERCUBAAN
MATEMATIK
UPSR**

ABD WAHAB BIN MUDA

**DISERTASI DIKEMUKAKAN INI UNTUK MEMENUHI
SEBAHAGIAN DARIPADA SYARAT MEMPEROLEHI
IJAZAH SARJANA PENDIDIKAN
(PENGUJIAN, PENGUKURAN DAN PENILAIAN)**

**FAKULTI SAINS KOGNITIF DAN PEMBANGUNAN MANUSIA
UNIVERSITI PENDIDIKAN SULTAN IDRIS**

2006

PENGAKUAN

Saya mengaku disertasi ini adalah hasil kerja saya sendiri kecuali nukilan dan ringkasan yang setiap satunya saya jelaskan sumbernya.

27 Mac 2006

ABD WAHAB BIN MUDA
2002100518

PENGHARGAAN

Bersyukur kepada Allah kerana dengan hidayah dan taufiqNya dapat jua penyelidikan ini disempurnakan. Saya ingin merakamkan setinggi-tinggi penghargaan dan terima kasih terutama kepada penyelia, Prof. Madya Dr. Hj. Bhasah Hj. Abu Bakar yang banyak memberi tunjuk ajar, bimbingan serta teguran yang bermanfaat dalam menyiapkan penyelidikan ini.

Penghargaan dan terima kasih juga saya tujukan kepada Prof. Dato' Dr. Abu Bakar Nordin dan Prof. Dr. Bachok Md.Taib yang bertanggungjawab menilai serta memberi teguran dan cadangan membina diperingkat cadangan penyelidikan serta kepada Dr. Quek Bong Cheang sebagai pemeriksa terhadap penulisan disertasi ini.

Terima kasih juga ditujukan khas kepada Bahagian Perancangan dan Penyelidikan Dasar Pendidikan, Jabatan Pendidikan Negeri Terengganu, dan semua Guru Besar dan penyelaras UPSR di sekolah-sekolah penyelidikan, diatas segala sokongan dan kerjasama yang diberikan.

Penghargaan dan terima kasih juga kepada semua kakitangan perpustakaan Universiti Pendidikan Sultan Idris, Universiti Malaya, Universiti Kebangsaan Malaysia dan Universiti Utara Malaysia yang turut memberi sokongan dalam menyempurnakan disertasi ini.

Tidak lupa juga ucapan terima kasih ditujukan khas kepada kakitangan Fakulti Sains Kognitif dan Pembangunan Manusia dan kakitangan pejabat pusat pengajian siswazah, yang banyak membantu sepanjang proses melengkapkan disertasi ini.

Akhir sekali terima kasih kepada semua individu yang membantu saya secara langsung atau tidak langsung dalam menyempurnakan penyelidikan ini.

ABD WAHAB BIN MUDA

ABSTRAK

Penyelidikan ini bertujuan untuk menyelidiki perbezaan pencapaian berdasarkan jantina di dalam peperiksaan percubaan matematik UPSR. Item-item dikumpulkan kepada 4 kumpulan berdasarkan topik pelajaran, aras taksonomi kognitif Bloom, penggunaan gambarajah dan penggunaan nama-nama orang. Sampel kajian terdiri daripada murid-murid darjah enam yang dipilih berdasarkan kategori pencapaian iaitu cemerlang, sederhana dan lemah. Setiap kategori mengandungi 100 pelajar lelaki dan 100 pelajar wanita dari 100 buah sekolah rendah di negeri Terengganu. Instrumen yang digunakan dalam penyelidikan ini adalah ujian matematik kertas 1 yang mengandungi 40 item aneka pilihan. Ujian-t (independent samples t-test) digunakan untuk menguji hipotesis pada aras keertian 5%. Dapatan penyelidikan menunjukkan kumpulan item yang berdasarkan topik pelajaran adalah signifikan pada semua kategori pencapaian. Kumpulan item yang berdasarkan aras taksonomi kognitif Bloom untuk kategori pencapaian sederhana juga didapati signifikan. Kumpulan item yang berdasarkan topik pelajaran paling ketara menunjukkan bias untuk kategori pencapaian sederhana dan lemah manakala kumpulan item yang menggunakan gambarajah dan nama-nama orang didapati tidak bias kepada mana-mana jantina untuk semua kategori pencapaian.

ABSTRACT

Differences in achievement based on gender in UPSR mathematic trial examination

The purpose of the research is to study the differences in achievement based on gender in UPSR mathematic trial examinations. The items are grouped into 4 groups which are based on the topics of lesson, Bloom cognitive taxonomy levels, diagrams and person names used. The samples consist of year six students which are selected based on their performance which have been categoried into excellent, medium and low. Each category has 100 girls and 100 boys from 100 primary schools in Terengganu. The instrument used was UPSR trial's examinations, mathematic paper 1 which contains 40 multiple choice items. The t-test (independent samples t-test) were used to test the hypotheses with significant level at 5 %. The results of the study indicated that the groups of items based on the topics of lesson are significant for all categories of achievers. Items based on Bloom cognitive taxonomy levels for the medium group also found to be significant. All groups of items using person's names and diagrams for all of the categories achievers are not significant. The groups of items based on the topics of lesson are highly biased for the medium and low categories achievers. The groups of items using person's names and diagrams indicated that there are no biased based on genders for all of the categories achievers.

KANDUNGAN

	HALAMAN
PENGAKUAN	ii
PENGHARGAAN	iii
ABSTRAK	iv
ABSTRACT	v
BAB 1	
PENDAHULUAN	
1.1 Pengenalan	1
1.2 Pernyataan masalah	11
1.3 Hipotesis penyelidikan	20
1.4 Soalan penyelidikan	23
1.5 Objektif penyelidikan	24
1.5.1 Objektif khusus	24
1.5.2 Objektif umum	27
1.6 Kepentingan penyelidikan	27
1.7 Batasan penyelidikan	31

BAB 2

TINJAUAN LITERATUR

2.1	Pengenalan	33
2.2	Tinjauan literatur	33
2.3	Rumusan	41

BAB 3

METODOLOGI PENYELIDIKAN

3.1	Pengenalan	43
3.2	Reka bentuk penyelidikan	44
3.3	Pembolehubah penyelidikan	44
3.4	Prosedur penyelidikan	45
3.5	Instrumen penyelidikan	46
3.6	Kerangka konseptual	50
3.7	Populasi dan persampelan subjek penyelidikan	53
3.8	Pungutan data	55
3.9	Analisis data	57
3.10	Jangka masa penyelidikan	58

BAB 4

DAPATAN PENYELIDIKAN

4.1	Pengenalan	60
4.2	Pengujian hipotesis	62
4.2.1	Pengujian hipotesis kumpulan cemerlang	62
4.2.1.1	Hipotesis no1 1 :	63
4.2.1.2	Hipotesis no1 2 :	67
4.2.1.3	Hipotesis no1 3 :	69
4.2.1.4	Hipotesis no1 4 :	70
4.2.2	Pengujian hipotesis kumpulan sederhana	73
4.2.2.1	Hipotesis no1 1 :	74
4.2.2.2	Hipotesis no1 2 :	79
4.2.2.3	Hipotesis no1 3 :	82
4.2.2.4	Hipotesis no1 4 :	84
4.2.3	Pengujian hipotesis kumpulan lemah	86
4.2.3.1	Hipotesis no1 1 :	87
4.2.3.2	Hipotesis no1 2 :	92
4.2.3.3	Hipotesis no1 3 :	94
4.2.3.4	Hipotesis no1 4 :	97
4.3	Rumusan hasil ujian hipotesis	99

BAB 5**PERBINCANGAN DAN KESIMPULAN**

5.1	Pengenalan	101
5.2	Hasil ujian hipotesis	102
5.2.1	Item bias jantina dalam instrumen penyelidikan	102
5.2.2	Item bias jantina berdasarkan topik pelajaran	106
5.2.3	Item bias jantina berdasarkan aras taksonomi kognitif Bloom	112
5.2.4	Item bias jantina berdasarkan penggunaan gambarajah	113
5.2.5	Item bias jantina berdasarkan penggunaan nama-nama orang	116
5.3	Rumusan	117
5.4	Implikasi dan aplikasi penyelidikan	118
5.5	Implikasi terhadap sistem pentaksiran pendidikan	120
5.6	Cadangan memperbaiki kualiti item	122
5.6.1	Bias isi kandungan	124
5.6.2	Bias dari sudut penggunaan bahasa	125
5.6.3	Bias dari sudut struktur dan format item	125
5.7	Cadangan	129
5.7.1	Penyelidik dibidang pengukuran pendidikan	129
5.7.2	Penggubal-penggubal item ujian disemua peringkat	130
5.7.3	Badan-badan pelaksana peperiksaan kebangsaan	132

5.8	Kesimpulan	133
-----	------------	-----

RUJUKAN	134
---------	-----

LAMPIRAN

Surat kebenaran menjalankan penyelidikan oleh Jabatan Pelajaran Terengganu

Instrumen penyelidikan

Contoh borang penyelidikan

SENARAI JADUAL

JADUAL	TAJUK	HALAMAN
Jadual 1.1	Analisis keputusan UPSR tahun 2000, 2001, 2002 dan 2003 sekolah rendah Terengganu	10
Jadual 1.2	Kemasukan, enrolmen dan lulusan universiti awam di Malaysia mengikut jantina bagi pelajar ijazah pertama	12
Jadual 1.3	Bilangan murid/pelajar dan guru mengikut jantina, tahun dan sekolah	17
Jadual 3.4	Bilangan sekolah rendah di negeri Terengganu sehingga 01 Januari 2003	54
Jadual 3.5	Gred, markah dan pencapaian murid sekolah rendah	55
Jadual 3.6	Ubahsuai gred, markah dan pencapaian murid	56
Jadual 4.7	Topik pelajaran dan nombor item yang terlibat serta bilangannya	64
Jadual 4.8	Hasil ujian hipotesis nol 1 kumpulan cemerlang	66
Jadual 4.9	Nombor item dan bilangannya berdasarkan aras taksonomi	68
Jadual 4.10	Hasil ujian hipotesis nol 2 kumpulan cemerlang	69
Jadual 4.11	Nombor item yang menggunakan gambarajah	70
Jadual 4.12	Hasil ujian hipotesis nol 3 kumpulan cemerlang	71
Jadual 4.13	Nombor-nombor item yang menggunakan nama-nama orang	72
Jadual 4.14	Hasil ujian hipotesis nol 4 kumpulan cemerlang	73
Jadual 4.15	Topik pelajaran dan nombor item yang terlibat serta bilangannya	75
Jadual 4.16	Hasil ujian hipotesis nol 1 kumpulan sederhana	78

Jadual 4.17	Nombor item berdasarkan aras taksonomi	80
Jadual 4.18	Hasil ujian hipotesis no1 2 kumpulan sederhana	82
Jadual 4.19	Nombor item yang menggunakan gambarajah	83
Jadual 4.20	Hasil ujian hipotesis no1 3 kumpulan sederhana	84
Jadual 4.21	Nombor item yang menggunakan nama-nama orang	85
Jadual 4.22	Hasil ujian hipotesis no1 4 kumpulan sederhana	86
Jadual 4.23	Topik pelajaran dan nombor item yang terlibat serta bilangannya	88
Jadual 4.24	Hasil ujian hipotesis no1 1 kumpulan lemah	91
Jadual 4.25	Nombor item berdasarkan aras taksonomi	93
Jadual 4.26	Hasil ujian hipotesis no1 2 kumpulan lemah	94
Jadual 4.27	Nombor item yang menggunakan gambarajah	95
Jadual 4.28	Hasil ujian hipotesis no1 3 kumpulan lemah	96
Jadual 4.29	Nombor item yang menggunakan nama-nama orang	98
Jadual 4.30	Hasil ujian hipotesis no1 4 kumpulan lemah	99
Jadual 4.31	Kumpulan item yang signifikan berdasarkan kumpulan pencapaian	100
Jadual 5.32	Hubungan kumpulan pencapaian, topik pelajaran dan jantina bias	109
Jadual 5.33	Min skor setiap aras taksonomi mengikut kumpulan pencapaian	113
Jadual 5.34	Min skor penggunaan gambarajah mengikut kumpulan pencapaian	115
Jadual 5.35	Min skor penggunaan nama-nama orang mengikut kumpulan pencapaian	117

SENARAI RAJAH

RAJAH	TAJUK	HALAMAN
Rajah 3.1	Kerangka konseptual menunjukkan langkah-langkah analisis item bias	52
Rajah 3.2	Carta aliran pemindahan data ke dalam borang penyelidikan	57
Rajah 5.3	Analisis item bias jantina dalam hierarki proses pembinaan instrumen ujian	128

BAB I

PENDAHULUAN

1.1 Pengenalan

Cara tunggal guru-guru dan pelajar-pelajar dapat mengetahui sama ada pembelajaran telah berlaku dalam bilik darjah adalah dengan menggunakan ujian dan penilaian. Ujian penting selain untuk menilai kemajuan pelajar dan kejayaan guru, juga mengesan kebolehan dan kelemahan pelajar serta menilai pembelajaran yang sedang berlaku supaya pelajar atau guru dapat mengetahui tindakan susulan yang seharusnya diambil. Kesemua pelajar akan mengalami penderitaan kerana kegagalan dan kenikmatan kerana kejayaan oleh proses pengujian. Keadaan ini akan lebih dideritai oleh pelajar yang gagal disebabkan oleh ujian yang bias (berat sebelah).

Ujian sepatutnya memberikan peluang yang sama kepada semua pelajar untuk memperagakan kebolehan dan pengetahuan mereka, namun kadangkala ujian-ujian yang dibina kerap kali mengandungi item-item yang bias tanpa disedari. Keadaan sebegini akan menyebabkan terdapat kumpulan-kumpulan pelajar yang berbeza pencapaiannya berdasarkan ujian tersebut walaupun mereka mempunyai kebolehan yang setara.

Ujian yang bias menjejaskan kesahan skor ujian dan mendatangkan kesan yang serius kepada kumpulan yang terlibat dan juga individu kerana kebarangkalian untuk memberi respons yang betul bukan hanya ditentukan oleh perkara yang diukur, tetapi juga oleh faktor-faktor yang khusus kepada kumpulan-kumpulan yang mengambil ujian, contohnya faktor perbezaan peluang dalam pembelajaran atau perbezaan sosial (Maller, 2001).

Bias boleh berlaku dalam berbagai bentuk. Ia mungkin bias jantina, budaya, etnik, agama atau kelas (Hambleton, R.K. dan Rodgers, H.J., 2000). Bias boleh berlaku jika item mempunyai kandungan atau bahasa yang berbeza penggunaannya oleh subkumpulan-subkumpulan tertentu yang mengambil ujian, atau jika struktur item atau formatnya yang berbeza kesukarannya untuk subkumpulan-subkumpulan yang mengambil ujian (Scheuneman, 1982a).

Bias dari sudut penggunaan bahasa dalam sesuatu item dikatakan wujud jika sesuatu item itu mengandungi perkataan yang berbeza makna atau kurang dikenali atau kosa kata yang sukar difahami terhadap mana-mana kumpulan yang mengambil ujian. Penggunaan bahasa dan kosa kata yang hanya khusus digunakan oleh

kumpulan-kumpulan tertentu juga akan mendatangkan kesukaran terhadap kumpulan yang lain.

Bias isi kandungan merujuk kepada item yang mempunyai isi kandungan yang berbeza atau kurang dikenali oleh kumpulan-kumpulan yang mengambil ujian sehingga menghasilkan jawapan betul atau salah dengan alasan yang salah, juga isi kandungan item yang dijawab oleh calon-calon yang mempunyai latar belakang pendidikan yang sama tetapi memberikan maklumbalas atau menghasilkan kemahiran yang tidak dijangkakan atau tidak ada kaitan dengan apa yang disoal.

Item yang bias dari segi format atau struktur boleh dikesan sekiranya terdapat petunjuk (*clues*) dalam item yang memberi kelebihan kepada satu kumpulan untuk mengatasi yang lain. Arahan-arahan ujian, item stem dan pengganggu yang tidak lengkap atau kabur atau mempunyai makna berbilang (*ambiguities*) turut menyumbang kepada bias.

Penggunaan bahasa, isi kandungan, struktur dan format item yang dibincangkan tadi sekiranya memberikan kelebihan kepada mana-mana kelompok jantina akan menyebabkan wujudnya bias jantina dalam item dan seterusnya menghasilkan ujian yang bias jantina.

Satu ujian dikatakan bias jantina jika lelaki dan perempuan yang mempunyai kebolehan yang sama arasnya cenderung mendapatkan skor yang berbeza (Childs, 1990). Begitu juga dengan item yang dijawab oleh pelajar. Sesuatu item itu dikatakan item yang bias jantina jika lelaki dan perempuan yang mempunyai kebolehan yang sama arasnya, cenderung untuk memberikan jawapan yang berbeza apabila menjawab item yang sama.

Kewujudan item bias boleh dikesan terhadap sesuatu ujian yang antara lainnya mempunyai ciri-ciri seperti penghasilan skor pencapaian yang berbeza oleh individu yang sama kebolehan tetapi datang dari berbeza jantina, etnik, budaya atau kumpulan-kumpulan agama.

Jika bias merujuk kepada ciri-ciri item itu sendiri, maka keadilan (fairness) pula merujuk kepada tujuan hasil ujian itu digunakan. Berlaku adil kepada semua calon yang terlibat dalam ujian merupakan salah satu dari prinsip pentaksiran pendidikan yang digariskan oleh National Center for Fair and Open Testing, Amerika Syarikat (2002). Menurut Luppescu (1993), bias berlaku apabila sesuatu item itu menunjukkan perbezaan aras kesukaran kepada satu kumpulan, jika dibandingkan dengan kumpulan lain yang mempunyai kebolehan yang sama. Atau dengan lain perkataan, item yang bias akan menyebabkan perbezaan prestasi antara kumpulan-kumpulan calon yang setara kebolehannya. Ini akan menimbulkan persoalan sama ada item yang sebegini rupa, adil atau tidak terhadap calon-calon yang mengambil ujian.

Jika sesuatu ujian itu terbukti bias, maka keputusan daripada ujian tersebut (skor atau gred) apabila digunakan oleh pelbagai pihak untuk menempatkan calon secara objektif untuk pelbagai tujuan, contohnya penentuan kedudukan di dalam kelas di sekolah, pengambilan ke sekolah terkawal, pemilihan murid-murid bagi menerima apa-apa anugerah, ganjaran, biasiswa, peluang melanjutkan pelajaran ke IPT atau peluang mendapatkan pekerjaan akan diragui keadilannya. Oleh yang demikian, pembina ujian perlu berusaha untuk menentukan ujian yang digunakan itu

tidak ‘cenderung’ kepada mana-mana kumpulan calon yang mengambil sesuatu peperiksaan.

Bias boleh dibahagikan kepada dua jenis iaitu *vernacular bias* dan *statistical bias*. Vernacular bias merujuk kepada keadaan dimana sesuatu ujian dianggap tidak adil manakala statistical bias merujuk kepada ciri-ciri psikometrik sesuatu ujian. Pendekatan yang lebih empirikal dalam menentukan item bias jantina boleh dibuat dengan mengkaji kesukaran item dengan memperingkatkan item berdasarkan kesukaran item merentasi jantina. Item yang menyebabkan kesukaran yang lebih terhadap sesuatu jantina untuk menjawabnya dianggap bias jantina.

Secara konsep, bias boleh wujud dari aspek pertalian diantara ujian dengan kriteria yang diuji atau sebagai item secara individu. Dari segi item, terjadinya bias boleh dinilai oleh pengadil (yang boleh mengenalpasti sesuatu item itu bias melalui cerapan) atau melalui penelitian ke atas ciri-ciri psikometriknnya (Cunningham, 1986).

Dalam sebarang penyiasatan bias (Hambleton, Ronald & Rodgers, Jane 1995), langkah pertama yang perlu dibuat ialah mengenalpasti item-item yang lebih menguntungkan atau memberi faedah atau memberi peluang yang lebih baik untuk menjawab dengan tepat kepada kumpulan-kumpulan tertentu atau terma-terma dalam item yang lebih memudahkan atau lebih ‘*familiar*’ kepada kumpulan-kumpulan tertentu (*designated subgroups of interest, DSI*).

Menurut Hambleton, Swaminathan dan Rodgers (1991), untuk mengatakan sesuatu item itu bias, kita haruslah mempunyai bukti empirikal bagi menyatakan dengan cekap bahawa sesuatu item itu bias. Oleh yang demikian untuk membezakan

antara bukti empirikal daripada rumusan, istilah Differential Item Functioning (DIF) daripada istilah bias atau Kefungsian Item Pelbagai (KIP) digunakan bagi menerangkan bukti empirikal yang didapati dalam penyelidikan bias. DIF dikatakan wujud apabila calon yang mempunyai kebolehan yang sama tetapi dari kumpulan yang berlainan (lelaki dan perempuan) mempunyai kebarangkalian yang berbeza dalam memilih jawapan yang betul bagi sesuatu item (Mazor Kanjee & Clauser, 1995).

Kaedah yang paling ringkas dan mudah dilakukan untuk menentukan bias ialah dengan membuat analisis item ke atas setiap item. Seorang ahli psikometrik boleh menentukan sama ada sesuatu item itu bias atau tidak, menggunakan pengalaman dan kepakarannya.

Di Malaysia, kita masih menggunakan Teori Ujian Klasik (Classical Test Theory -CTT) dalam membina ujian yang berdasarkan kepada kesukaran dan diskriminasi item. Kedua-dua statistik ini adalah bergantung-kumpulan dimana ujian-ujian yang dibina dipraujikan terlebih dahulu kepada kumpulan-kumpulan yang dianggap representatif atau yang hampir sama kebolehannya dengan sampel yang bakal digunakan untuk memastikan ujian yang dibina berkualiti tinggi.

Dalam penggunaan model CTT, ciri-ciri murid dan ciri-ciri ujian tidak boleh dipisahkan. Ciri murid yang dimaksudkan ialah kebolehan murid yang dinyatakan sebagai markah yang didapati dalam ujian (skor cerapan). Ini bermaksud kebolehan murid diterangkan berdasarkan skor sesuatu ujian. Sekiranya ujian yang diberikan itu sukar, ramai murid tidak dapat menjawab item berkenaan, maka murid dianggap berkebolehan rendah, begitu jugalah sebaliknya. Walaupun kaedah CTT nampak

mudah tetapi ternyata kaedah ini mempunyai beberapa kelemahannya. Diantara kelemahan-kelemahan itu ialah:

- (i) Ciri-ciri itemnya bergantung-kumpulan.
- (ii) Skor yang menghuraikan kecekapan murid bergantung-ujian.
- (iii) Model tidak dapat menjelaskan hingga keparas item.
- (iv) Model memerlukan ujian selari untuk mentaksir kebolehpercayaan.
- (v) Model tidak dapat memberikan pengukuran yang tepat pada setiap skor kebolehan.

Oleh kerana CTT lebih berorientasikan ujian daripada item, maka CTT tidak dapat meramalkan prestasi murid apabila sesuatu item diberikan untuk dijawab. Sebagai alternatif terhadap kesukaran yang dihadapi menggunakan CTT, Item Response Theory (IRT) yang merupakan satu teori moden dalam bidang pengukuran pendidikan diperkenalkan oleh ahli-ahli psikometrik. Model pengukuran mental ini adalah pambinaan dari model CTT. IRT juga disebut sebagai Teori Trait Pendaman atau Teori Ciri Keluk Item (Item Characteristic Curve – ICC). IRT merupakan satu keluarga model matematik yang menghuraikan perlakuan item dalam sesuatu ujian

berdasarkan fungsi respons item (FRI). Kini, IRT kerap digunakan oleh para penyelidik untuk mengesan item bias.

Tujuan utama melakukan penyelidikan item bias ialah untuk menentukan sama ada skor ujian dipengaruhi oleh sumber variasi yang berbeza dalam sampel yang berbeza. Jika didapati skor ujian itu benar-benar dipengaruhi oleh sumber varian yang sama, maka kita perlu melihat sama ada sumber-sumber itu memberi kelebihan kepada sesuatu sampel pelajar.

Kewujudan DIF dalam kertas ujian boleh dikenalpasti secara membandingkan fungsi ciri item daripada kedua-dua kumpulan berkenaan (lelaki dan perempuan). Ini boleh dilakukan dengan membandingkan parameter ICC bagi kumpulan berkenaan, DIF wujud sekiranya ICC kedua-dua kumpulan itu berbeza. Selain daripada itu, ia boleh dibuat secara membandingkan ICC bagi kedua-dua kumpulan itu dan menghitung keluasan yang wujud antara kedua-dua keluk itu. Jika keluasan antara kedua-dua ICC itu sifar, ini menunjukkan ICC itu bertindih, maka tidak wujud sebarang DIF bagi item berkenaan. Ini bermakna item berkenaan tidak bias.

Penyelidikan Kefungsian Item Pelbagai ialah satu prosedur yang bertujuan mengenalpasti item yang berfungsi secara berbeza bila ditadbirkan kepada sekumpulan calon yang mempunyai perbezaan ciri atau latar belakang tetapi setara kebolehannya. Menurut Holland dan Wainer (1992), item yang memperlihatkan indeks KIP atau DIF ini tidak membuktikan bahawa ia cenderung (bias). Prosedur ini hanya dapat membantu mengenal pasti item yang mungkin bersifat tidak adil dan perlu dikaji semula untuk penambahbaikan atau digugurkan. Matlamat prosedur ini

ialah untuk mengurangkan atau menghapuskan item ujian yang berpotensi menyebabkan wujudnya bias.

Satu kaedah yang lebih mudah untuk mengesan item bias ialah menggunakan ujian khi kuasa dua (ujian tak berparameter), iaitu dengan melakukan ujian kebebasan (test for independent) terhadap pembolehubah-pembolehubah yang berkaitan. Ujian kebebasan dilakukan bagi mencari adakah terdapat hubungan antara jantina murid dan skor mereka bagi setiap item dengan syarat gred atau aras pencapaian pelajar (lelaki dan perempuan) adalah setara. Untuk menguji hipotesis nol, nilai statistik ujian dikira dengan mencari kadar nilai kekerapan dicerap berbanding dengan nilai kekerapan yang dijangka (data menggunakan skala nominal). Jika nilai p kurang dari 0.05 hipotesis nol ditolak maka keputusannya ialah tidak terdapat hubungan antara jantina murid dengan skor mereka (tiada bias).

Satu lagi kaedah ialah menggunakan ujian t (independent-sample t test) iaitu menguji kewujudan perbezaan antara min satu pembolehubah bagi dua kumpulan sampel yang tidak saling bergantung. Untuk menguji hipotesis nol, nilai statistik ujian dikira dengan mencari perbezaan min skor kumpulan pelajar lelaki dan perempuan bagi satu-satu atau sekumpulan item mengikut kriteria yang telah ditetapkan. Penyelidik akan menggunakan kedua-dua kaedah ini atau salah satu daripadanya dimana yang lebih sesuai untuk mengesan item bias jantina dalam ujian matematik kertas 1, Ujian Percubaan UPSR negeri Terengganu.

Keputusan UPSR sekolah-sekolah di negeri Terengganu mencatatkan peningkatan pencapaian 5A selama tiga tahun berturut-turut dan mengungguli negeri-negeri lain dalam negara. Namun apa yang jelas ialah, bilangan murid

perempuan yang mencapai A dalam semua mata pelajaran semakin bertambah malah semakin jauh meninggalkan murid lelaki, sedangkan bilangan calon yang mengambil peperiksaan untuk kedua-dua jantina ini hampir sama. Pada tahun 2003, bilangan calon lelaki adalah 12,928 dan wanita 12,257 iaitu calon lelaki melebihi calon wanita seramai 671 orang, sedangkan keputusan UPSR menunjukkan pencapaian semua A didahului oleh pelajar wanita. Daripada jumlah pelajar yang mendapat semua A seramai 2,710, seramai 1019 adalah lelaki dan 1691 adalah wanita. Keputusan yang diperolehi ini menunjukkan calon perempuan yang mendapat A dalam semua mata pelajaran melebihi calon lelaki seramai 672 orang.

Jadual 1.1

Analisis keputusan UPSR tahun 2000, 2001, 2002 dan 2003 sekolah rendah Terengganu.

TAHUN	CALON	PENCAPAIAN 5A/7A	%
2000	24546	1482	6.04
2001	24147	2223	9.21
2002	24966	2498	10.01
2003	25185	2710	10.76