

KEBOLEHPERCAYAAN PENILAIAN GURU TERHADAP
KERJA KURSUS MATA PELAJARAN KEMAHIRAN HIDUP
PROGRAM PERALIHAN

oleh

SHUHAILI BIN ABD RAHIM

Tesis diserahkan untuk memenuhi
keperluan bagi Ijazah Sarjana Pendidikan

Mei 1997

MIKROFIS

5240

UCAPAN PENGHARGAAN

Bismillahirrahmanirrahim

Saya memanjatkan setinggi-tinggi kesyukuran ke hadrat Allah SWT kerana dengan taufiq dan hidayahNya membolehkan saya menyiapkan tesis ini dengan jayanya. Ucapan penghargaan ini saya rakamkan khas untuk penyelia utama iaitu Dr Kamaruddin bin Yaakub kerana tidak kenal jemu memberikan dorongan dan bimbingan sehingga tesis ini dapat disiapkan sepenuhnya. Ucapan ini juga dirakamkan untuk semua staf PPIP, IPS dan Perpustakaan kerana menyediakan segala kemudahan pengajian serta Universiti Sains Malaysia kerana bermurah hati memberikan peruntukan gran penyelidikan.

Ucapan terima kasih kepada Bahagian Biasiswa Kementerian Pendidikan Malaysia kerana memberikan peluang kepada saya untuk mengikuti pengajian ini dengan biasiswa. Begitu juga kepada EPRD, JPN Kedah, pengetua, guru dan pelajar di sekolah-sekolah yang terlibat kerana memberikan kerjasama sepenuhnya dalam menyediakan sampel dan data kajian. Ucapan khas dirakamkan untuk ibubapa, isteri, anak-anak, rakan dan sahabat karib saya Encik Hashim Che Mat kerana sentiasa memberikan sokongan moral yang kuat. Segala jasa baik anda sekalian sesungguhnya tidak akan saya lupai. Semoga Allah SWT membalasnya dengan memberikan pahala yang setimpal. Amin. Wassalam

JADUAL KANDUNGAN

Muka Surat

UCAPAN PENGHARGAAN	11
SENARAI JADUAL DAN RAJAH	vii
SENARAI SINGKATAN	x
ABSTRAK	xi
ABSTRACT	xiii
BAB SATU: MASALAH DAN KEDUDUKANNYA DALAM KAJIAN	
1.1 Latar Belakang Kajian -----	1
1.2 Tujuan Kajian -----	7
1.3 Pernyataan Masalah -----	9
1.4 Hipotesis Kajian -----	10
1.5 Definisi Istilah Operasional -----	12
1.6 Batasan Kajian -----	16
1.7 Andalan Kajian -----	16
BAB DUA: KAJIAN KARYA YANG BERKAITAN	
2.1 Pengenalan -----	18
2.2 Kebolehpercayaan Markah Peperiksaan -----	18
2.3 Unsur Harapan Guru -----	30
2.4 Bentuk Interaksi Guru Dengan Pelajar -----	35
2.5 Kecekapan Guru Sebagai Pemeriksa -----	38
2.6 Rumusan Bacaan Karya Yang Berkaitan -----	42
BAB TIGA: KAEDAH KAJIAN	
3.1 Pengenalan -----	45
3.2 Tempat Kajian -----	45

3.3	Sampel Kajian -----	45
3.4	Pendekatan Kajian -----	51
3.5	Alatukur -----	56
3.6	Pengumpulan Data -----	57
3.7	Analisis Data -----	58

BAB EMPAT: KEPUTUSAN KAJIAN

4.1	Pengenalan -----	65
4.2	Analisis Deskriptif -----	65
4.2.1	Taburan Kekkerapan Markah -----	65
4.2.2	Ukuran Kecenderungan Memusat -----	72
4.2.3	Ukuran Kebolehubahan -----	74
4.2.4	Analisis Markah Selisih -----	77
4.3	Analisis Statistik Inferens -----	85
4.3.1	Kebolehpercayaan Ketekalan Dalam ---	85
4.3.2	Kebolehpercayaan Intra-Pemeriksa ---	92
4.3.3	Kebolehpercayaan Inter-Pemeriksa ---	96
4.3.4	Tafsiran Keputusan Bagi Keboleh- percayaan Markah Diperolehi -----	107
4.3.5	Pengujian Hipotesis -----	112
4.4	Rumusan Dapatan Kajian -----	124

BAB LIMA: PERBINCANGAN DAN IMPLIKASI KAJIAN

5.1	Pengenalan -----	127
5.2	Perbincangan -----	127
(1)	Keberkesanan Pemeriksa Melaksanakan Aktiviti Pemarkahan Projek Kerja Kursus Praktikal -----	127

(2)	Konflik Peranan Yang Dialami Oleh Guru Yang Bertugas Sebagai Pemeriksa -----	134
(3)	Kesan Pengulangan Aktiviti Pemarkahan Oleh Pemeriksa Yang Sama -----	136
(4)	Ketelitian Pemeriksa Menggunakan Skema Pemarkahan -----	137
(5)	Persepsi Pemeriksa Terhadap Beban Tugas -----	140
5.3	Implikasi Kajian -----	142
5.4	Cadangan Penyelidikan Lanjutan -----	148
BIBLIOGRAFI -----		148
LAMPIRAN		
A	Lukisan Kerja Projek Bekas Makanan Haiwan ---	159
B	Skema Pemarkahan Projek Bekas Makanan Haiwan -----	161
C	Borang Pemarkahan Projek Bekas Makanan Haiwan -----	162
D	Arahan Pemeriksaan kepada Guru/Penilai/Pemeriksa -----	163
E	Langkah-Langkah Membuat Projek -----	164
F	Markah Diperolehi bagi Projek Bekas Makanan Haiwan MPKHPP Pilihan B bagi Sekolah/Pemeriksa -----	166
G	Jadual Taburan Kekerapan Markah Projek Bekas Makanan Haiwan MPKHPP Pilihan B bagi Sekolah/Pemeriksa -----	174
H	Markah Min Projek Bekas Makanan Haiwan bagi Markah MPS, MPD, MPP dan MPK -----	198
I	Markah Selisih antara Markah MPS, MPD, MPP dan MPK bagi Sekolah/Pemeriksa -----	200

J	Jadual Matriks Koefisien Alpha bagi Item-Item (25 Item) Projek Bekas Makanan Haiwan bagi Markah MPS, MPD, MPP Dan MPK -----	208
K	Julat Markah Benar mengikut Teori Ujian Klasik bagi Markah Projek Bekas Makanan Haiwan bagi Bekolah/Pemeriksa -----	212

SENARAI JADUAL DAN RAJAH

Jadual	Tajuk	Muka Surat
1	Bilangan Pelajar Tingkatan 3 tahun 1992 yang telah membuat Projek Bekas Makanan Haiwan -----	49
2	Nombor Kod Kawalan bagi Projek Bekas Makanan Haiwan -----	59
3	Bilangan Peruntukan Projek kepada Pemeriksa-Penyetara dan Pemeriksa Khas -----	60
4	Taburan Kekerapan Markah MPS, MPD, MPP dan MPK -----	66
5	Kedudukan Nilai Pencong bagi Taburan Kekerapan Markah MPS, MPD, MPP dan MPK -----	70
6	Min, Median dan Mod bagi Markah MPS, MPD, MPP dan MPK -----	71
7	Markah Min bagi Markah MPS, MPD, MPP dan MPK -----	73
8	Selisih Markah Min bagi Markah MPS, MPD, MPP dan MPK -----	75
9	Sisihan Piawai bagi Markah MPS, MPD, MPP dan MPK -----	76
10	Koefisien Ubahan bagi Markah MPS, MPD, MPP dan MPK -----	78
11	Rumusan Analisis Markah Selisih bagi Item-Item (25 Item) (n = 256) -----	80
12	Analisis Markah Selisih antara Markah MPS, MPD, MPP dengan Markah MPK (n = 256) -----	82
13	Analisis Markah Selisih antara Markah MPS, MPD, MPP dengan Markah MPK (n = 32) -----	84
14	Anggaran Nilai Koefisien Alpha bagi Item-Item (25 Item) bagi Markah MPS, MPD, MPP dan MPK (n = 256) -----	86

15	Anggaran Nilai Koefisien Alpha bagi Item-Item (25 Item) bagi Markah MPS, MPD, MPP dan MPK (n = 32) -----	87
16	Anggaran Nilai Koefisien Korelasi Intra-Pemeriksa antara Markah MPS dengan Markah MPD -----	93
17	Anggaran Nilai Koefisien Penentuan dan Koefisien Kecekapan Ramalan bagi Koefisien Korelasi Intra-Pemeriksa antara Markah MPS dengan Markah MPD -----	95
18	Anggaran Nilai Koefisien Korelasi Inter-Pemeriksa antara Markah MPS dengan Markah MPP -----	97
19	Anggaran Nilai Koefisien Penentuan dan Koefisien Kecekapan Ramalan bagi Koefisien Korelasi Inter-Pemeriksa antara Markah MPS dengan Markah MPP -----	100
20	Anggaran Nilai Koefisien Korelasi Inter-Pemeriksa antara Markah MPS dengan Markah MPK -----	102
21	Anggaran Nilai Koefisien Penentuan dan Koefisien Kecekapan Ramalan bagi Koefisien Korelasi Inter-Pemeriksa antara Markah MPS dengan Markah MPK -----	104
22	Anggaran Nilai Koefisien Korelasi Inter-Pemeriksa antara Markah MPP dengan Markah MPK -----	106
23	Anggaran Nilai Koefisien Penentuan dan Koefisien Kecekapan Ramalan bagi Koefisien Korelasi Inter-Pemeriksa antara Markah MPP dengan Markah MPK -----	108
24	Anggaran Nilai Ralat Piawai Pengukuran bagi pemarkahan Intra-Pemeriksa dan Inter-Pemeriksa (n = 256) -----	109
25	Anggaran Nilai Ralat Piawai Pengukuran bagi pemarkahan Intra-Pemeriksa dan Inter-Pemeriksa (n = 32) -----	113
26	Ujian-t bagi pemarkahan Intra-Pemeriksa antara Markah MPS dengan Markah MPD -----	117

27	Ujian-t bagi pemarkahan Inter-Pemeriksa antara Markah MPS dengan Markah MPP -----	119
28	Ujian-t bagi pemarkahan Inter-Pemeriksa antara Markah MPD dengan Markah MPP -----	120
29	Ujian-t bagi pemarkahan Inter-Pemeriksa antara Markah MPS dengan Markah MPK -----	122
30	Ujian-t bagi pemarkahan Inter-Pemeriksa antara Markah MPP dengan Markah MPK -----	123
31	Rumusan Markah Selisih antara Markah MPS, MPD, MPP dan MPK -----	133

Rajah

1	Rangkaian Unsur Harapan -----	34
2	Peringkat-Peringkat Pemilihan Sampel Kajian -----	46
3	Taburan Kekerapan Markah MPS, MPD, MPP dan MPK (carta bar) -----	67
4	Taburan Kekerapan Markah MPS, MPD, MPP dan MPK (carta garisan) -----	69

SENARAI SINGKATAN

- k - kekerapan
- MPS - Markah Pemeriksa bagi Aktiviti Pemarkahan pada kali yang Pertama
- MPD - Markah Pemeriksa bagi Aktiviti Pemarkahan pada kali yang Kedua
- MPP - Markah Pemeriksa-Penyetara
- MPK - Markah Pemeriksa Khas
- MPKHPP - Mata Pelajaran Kemahiran Hidup Program Peralihan
- n - bilangan sampel projek
- P - Pemeriksa
- PBMH - Projek Bekas Makanan Haiwan
- Pilihan B - pilihan Perdagangan & Keusahawanan
- SP - Sisihan Piawai
- RPP - Ralat Piawai Pengukuran

ABSTRAK

Kajian ini telah dijalankan di sekolah-sekolah menengah dalam Negeri Kedah. Sekolah-sekolah tersebut ialah sekolah yang menawarkan MPKHPP pilihan B. Sampel kajian terdiri daripada 8 buah sekolah, 8 orang pemeriksa, 4 orang pemeriksa-penyetara, seorang pemeriksa khas dan 256 buah projek bekas makanan haiwan yang telah dibuat oleh pelajar tingkatan 3 tahun 1992. Kajian ini menggunakan dua alatukur iaitu lukisan kerja projek bekas makanan haiwan dan skema pemarkahan. Alatukur ini digunakan untuk mendapatkan markah diperolehi. Markah diperolehi seterusnya dianalisis bagi tujuan untuk meninjau kedudukannya dari segi kebolehpercayaan ketekalan dalam, kebolehpercayaan intra-pemeriksa dan kebolehpercayaan inter-pemeriksa.

Keputusan kajian telah menunjukkan bahawa terdapat satu pertalian positif yang kukuh dan boleh dipercayai bagi kebolehpercayaan ketekalan dalam. Nilai alpha yang diperolehi ialah .8464 bagi Markah MPS, .8719 bagi Markah MPD, .8242 bagi Markah MPP dan .7483 bagi Markah MPK. Begitu juga dengan kedudukan bagi kebolehpercayaan intra-pemeriksa. Nilai koefisien korelasi yang diperolehi

ialah .9234. Lima daripada lapan orang pemeriksa telah menunjukkan nilai alpha melebihi daripada .90.

Keputusan bagi kebolehpercayaan inter-pemeriksa pula telah menunjukkan satu kedudukan nilai koefisien korelasi yang menjurus ke arah nilai yang rendah. Nilai korelasi yang diperolehi antara Markah MPS dengan Markah MPP ialah .5446, antara Markah MPS dengan Markah MPK ialah .3179 dan antara Markah MPP dengan Markah MPK ialah .2787. Secara keseluruhannya nilai ini adalah tidak memuaskan. Seterusnya pengujian hipotesis nol dengan menggunakan kaedah Ujian-t telah menunjukkan keputusan di mana keempat-empat hipotesis nol yang diuji pada aras signifikan $p < .05$ adalah ditolak. Sehubungan dengan itu maka dapatan kajian ini nampaknya adalah tekal dengan dapatan lain-lain kajian berhubung dengan kebolehpercayaan intra-pemeriksa dan kebolehpercayaan inter-pemeriksa.

ABSTRAK

Kajian ini telah dijalankan di sekolah-sekolah menengah dalam Negeri Kedah. Sekolah-sekolah tersebut ialah sekolah yang menawarkan MPKHPP pilihan B. Sampel kajian terdiri daripada 8 buah sekolah, 8 orang pemeriksa, 4 orang pemeriksa-penyetara, seorang pemeriksa khas dan 256 buah projek bekas makanan haiwan yang telah dibuat oleh pelajar tingkatan 3 tahun 1992. Kajian ini menggunakan dua alatukur iaitu lukisan kerja projek bekas makanan haiwan dan skema pemarkahan. Alatukur ini digunakan untuk mendapatkan markah diperolehi. Markah diperolehi seterusnya dianalisis bagi tujuan untuk meninjau kedudukannya dari segi kebolehpercayaan ketekalan dalam, kebolehpercayaan intra-pemeriksa dan kebolehpercayaan inter-pemeriksa.

Keputusan kajian telah menunjukkan bahawa terdapat satu pertalian positif yang kukuh dan boleh dipercayai bagi kebolehpercayaan ketekalan dalam. Nilai alpha yang diperolehi ialah .8464 bagi Markah MPS, .8719 bagi Markah MPD, .8242 bagi Markah MPP dan .7483 bagi Markah MPK. Begitu juga dengan kedudukan bagi kebolehpercayaan intra-pemeriksa. Nilai koefisien korelasi yang diperolehi

correlation index is .9234. Five out of eight examiners have shown an alpha value exceeding .90.

The inter-examiners reliability showed a coefficient correlation index bias towards the lower values. The coefficient correlation index are as follows: between MPS scores and MPP scores is .5446, between MPS scores and MPK scores is .3179 and between MPP scores and MPK scores is .2787. On the whole the findings are indeed not very satisfactory. Further, the null hypotheses that have been tested by utilizing t-test method showed that the four null hypotheses can be rejected at $p < .05$ significant level. In relation to this, the findings of this study have consistently supported the assumption of the other studies carried out in relation to the intra-examiners and inter-examiners reliability.

MASALAH DAN KEDUDUKANNYA DALAM KAJIAN

1.1 Latar Belakang Kajian

Tindakan positif Kementerian Pendidikan Malaysia menyerahkan tugas menilai projek kerja kursus praktikal kepada guru yang mengajar pelajarannya sendiri di bawah konsep Penilaian Kendalian Sekolah Menengah (PKSM) bagi mata pelajaran Kemahiran Hidup, Pendidikan Islam, Sejarah dan Geografi telah menarik perhatian ramai ibu bapa (Berita Harian, 1991, ms. 10; Berita Harian, 1991, ms. 12). Mereka telah menyuarakan pandangan bahawa sikap seseorang guru yang melaksanakan tugas menilai projek kerja kursus praktikal yang dibuat oleh pelajarannya sendiri mungkin boleh menyebabkan markah yang diberikan menjadi tidak adil dan kurang boleh dipercayai. Guru mungkin memberikan lebih markah kepada pelajar yang disukainya dan sebaliknya memberikan kurang markah kepada pelajar yang menjadi bukan kesukaan guru. Ini adalah kerana pelajar ini adalah satu golongan masyarakat yang berbeza dari segi keturunan, latar belakang sosio-ekonomi, jantina dan pencapaian.

Siti Hawa Ahmad (1989) dalam kertas kerja yang bertajuk 'The Role of School-Based Assessment: The Malaysian Experience' telah menyatakan bahawa pelaksanaan Penilaian Kendalian Sekolah (PKS) yang mula dilaksanakan pada tahun 1987 adalah bertujuan untuk memberikan peluang seluas-luasnya kepada guru dan sekolah untuk meningkatkan perkembangan keseluruhan kemajuan pelajaran dan juga untuk

membantu mengurangkan tekanan yang berlebihan semata-mata terhadap peperiksaan. Kenyataan ini nampaknya dapat disokong dengan pendapat yang telah dikemukakan oleh Cohen dan Deale (1977) dan Nduka Okoh (1983). Mereka telah menyebut bahawa guru kelas atau guru mata pelajaran merupakan orang yang berada pada satu kedudukan yang paling sesuai untuk menilai pelajarannya sendiri. Ini kerana pada asasnya guru itu adalah merupakan orang yang benar-benar mengetahui dan memahami tentang kebolehan pelajarannya. Guru itu juga merupakan orang yang mempunyai kecakapan profesional dalam bidang penilaian. Oleh yang demikian maka kedudukan ini dapat meletakkan kaum guru sebagai orang yang mampu untuk membuat penilaian dengan itu.

Mata pelajaran Kemahiran Hidup Program Peralihan (MPKHPP) mula dilaksanakan oleh Kementerian Pendidikan Malaysia pada tahun 1989 (Pusat Perkembangan Kurikulum, 1989). Mata pelajaran ini menawarkan empat pilihan iaitu pilihan A (Seni Perusahaan), pilihan B (Perdagangan dan Keusahawanan), pilihan C (Ekonomi Rumahtangga) dan pilihan D (Sains Pertanian). Kajian ini telah memilih MPKHPP pilihan B untuk meninjau kerja-kerja penilaian yang dilaksanakan oleh guru. Penilaian bagi MPKHPP pilihan B dilaksanakan dalam dua cara iaitu ujian bertulis dan projek kerja kursus praktikal. Ujian bertulis dikelolakan sendiri oleh Lembaga Peperiksaan. Penilaian projek kerja kursus praktikal pula telah diserahkan kepada guru sekolah. Guru sekolah melaksanakan penilaian projek kerja kursus praktikal ini berpandukan kepada beberapa arahan

penilaian yang telah dikeluarkan oleh Lembaga Peperiksaan.

Masyarakat secara umumnya bersetuju bahawa markah-markah diperolehi yang telah diberikan oleh seseorang guru atau pemeriksa kepada kertas-kertas peperiksaan objektif, teori dan kerja kursus praktikal bagi mata pelajaran yang telah diduduki oleh seseorang calon dalam sesuatu peperiksaan mempunyai nilai yang amat bermakna kepada calon tersebut. Markah-markah diperolehi itu boleh menjadi satu bukti kukuh tentang tahap pencapaian seseorang calon peperiksaan meskipun ia tidaklah boleh menjadi satu-satunya ukuran pencapaian yang mutlak (Macquarie University, 1971). Kebiasaannya calon akan menggunakan markah-markah diperolehi itu untuk meneruskan pelajaran ke peringkat yang lebih tinggi ataupun menggunakannya untuk memohon apa-apa pekerjaan yang sesuai.

Seterusnya kalau dipandang dari sudut kegunaannya kepada guru atau pegawai pendidikan maka manfaatnya akan tersebar lebih luas meliputi semua pelajar dan juga sistem pendidikan. Guru biasanya memanfaatkan markah-markah diperolehi itu sebagai satu maklumbalas bagi mengukur ketakat manakah pengetahuan, kefahaman dan kemahiran telah dapat dikuasai oleh pelajarnya. Markah-markah diperolehi itu boleh dijadikan sebagai asas untuk mengelaskan pelajarnya ke dalam beberapa kumpulan pencapaian tertentu (Hartog, 1958; Davis, 1964), untuk mengenalpasti apakah kekuatan dan kelemahan pelajar agar teknik pengajaran dapat diubahsuai (Spear, 1989) ataupun untuk memantapkan

plawaian kerja (Rowntree, 1987). Pegawai pendidikan pula akan memanfaatkan markah-markah diperolehi dalam bentuk merancang plan pembangunan pendidikan seperti kajian keciciran, kurikulum dan seterusnya sebagai satu ukuran penting mengenai akauntabiliti guru (Lembaga Peperiksaan, 1983) dan tahap keberkesanan sesebuah sekolah (Hartog, 1958).

Berdasarkan kepada besarnya kegunaan dan kuatnya pengaruh markah-markah diperolehi bagi sesuatu keputusan peperiksaan maka matlamat utama yang diharapkan dari markah-markah diperolehi itu ialah kejituan pengukurannya (Skurnik & Nuttall, 1968). Setakat manakah markah-markah diperolehi yang telah diberikan oleh seseorang pemeriksa itu boleh dianggap jitu? Ramai penyelidik telah merujuk persoalan kejituan markah-markah diperolehi ini melalui konsep kebolehpercayaan markah. Kata mereka, markah diperolehi yang jitu adalah markah-markah yang boleh dipercayai (Burt, 1945; Wiseman, 1949; Adams, 1964; Sulliksen, 1967; Thondike, 1971; Ary, et.al., 1972; Nuttall & Willmott, 1972), markah-markah yang stabil, bolehharap dan bolehtelah. (Kerlinger, 1973) dan paling sedikit mengandungi ralat pengukuran (Stanley, 1971).

Mengapakah markah-markah diperolehi yang telah diberikan oleh seseorang pemeriksa itu boleh dipercayai atau tidak boleh dipercayai? Berhubung dengan soalan ini Burt (1945) dan Willmott & Nuttall (1975) berpendapat bahawa markah-markah diperolehi menjadi tidak boleh dipercayai kerana berpunca dari masalah variasi dalam markah-markah itu. Variasi itu boleh berlaku meskipun

seorang pemeriksa yang sama memeriksa kertas peperiksaan secara berulang kali bagi calon yang sama dan kertas peperiksaan yang sama. Masalah variasi ini boleh dilihat berlaku dengan lebih jelas apabila beberapa pemeriksa yang berlainan memeriksa kertas peperiksaan yang sama dan bagi calon yang sama. Manakala markah-markah diperolehi yang boleh dipercayai tidaklah menjadi masalah.

Berhubung dengan masalah ketidakbolehpercayaan markah-markah diperolehi itu maka ramai penyelidik yang berminat dalam permasalahan ini telah mencadangkan berbagai pendekatan kuantitatif untuk menanganinya dari aspek-aspek yang mereka dipercayai boleh membantu menghurai dan melegakan masalah tersebut. Antara pendekatan-pendekatan yang telah dicadangkan itu ialah teori ujian klasik (Lord & Novick, 1968 (dalam Willmott & Nuttall, 1975)), kaedah penyetaraan markah (Backhouse, 1972), kaedah analisis faktor atau analisis varians (Burt, 1945), kaedah kebolehpercayaan ketekalan dalam (Willmott & Nuttall, 1972; 1975), kaedah ralat piawai pengukuran (Skurnik & Nuttall, 1968; Ary, et. al., 1972) dan kaedah koefisien korelasi hasil darab momen Pearson (Leedy, 1980). Persoalannya, teknik manakah yang paling berkesan untuk menangani masalah itu bergantunglah kepada kesesuaiannya dari segi bagaimanakah markah-markah diperolehi itu dikaji. Ini adalah kerana setiap teknik itu ada kekuatan dan kelemahan masing-masing.

Selain daripada menggunakan pendekatan kuantitatif untuk menangani masalah ketidakbolehpercayaan markah

perolehi maka terdapat juga penyidik yang mengabaikan dengan menggunakan pendekatan kualitatif. Misalnya, masalah ketidakbolehppercayaan markah diperolehi bagi jawapan kertas peperiksaan berbentuk esei, mereka telah menunjukkan bahawa tulisan yang cantik, mudah dibaca dan gaya persembahan jawapan yang menarik telah menyebabkan pemeriksa bersikap memberikan lebih markah berbanding dengan keadaan sebaliknya (Britton, et. al., 1966; Nuttall & Willmott, 1972; Branthwaite, et. al., 1981). Bagi memarkahan kerja kursus praktikal yang melibatkan proses penyediaan bahan, penilaian dan penyetaraan markah pula telah menimbulkan masalah yang lain iaitu kesukaran pemeriksa untuk memperuntukkan masa yang mencukupi untuk melaksanakan penilaian (Cohen & Deale, 1977; Selkirk, 1988). Seterusnya calon adalah bernasib baik jika minat pemeriksa sewaktu melaksanakan pemeriksaan adalah tinggi (Nuttall & Willmott, 1972; Townsend, et. al., 1989). Walau bagaimanapun keadaan yang paling malang sekali boleh berlaku jika pemeriksa memberikan markah tanpa membaca skrip jawapan (Britton, et. al., 1966), kurang pengalaman sebagai pemeriksa (Burt, 1945) dan ketidakcekapan pemeriksa secara peribadi (Wood & Quinn, 1976; Branthwaite, et. al., 1981; Jasman, 1989). Oleh yang demikian maka persoalan untuk mendapatkan markah-markah diperolehi yang boleh dipercayai akan menjadi lebih menepati keperluan jika kedua-dua pendekatan itu diambil kira bagi setiap kajian yang hendak dijalankan.

Tujuan Kajian

Kajian ini telah dijalankan dengan tujuan untuk meninjau keberkesanan melaksanakan pemarkahan pada projek kursus praktikal. Adakah markah-markah diperolehi telah diberikan pada projek kerja kursus praktikal oleh guru sebagai pemeriksa kepada projek yang telah dibuat oleh pelajar mereka sendiri boleh mewakili kebolehan sebenar pelajar itu?

Seseorang guru yang menjalankan tugas sebagai pemeriksa kertas jawapan atau projek kerja kursus praktikal kerap kali tersepit dalam pelbagai konflik dan tekanan. Mereka bukan sahaja tertekan dengan pelbagai kriteria pemarkahan yang diperlukan untuk menghasilkan markah diperolehi yang jitu bahkan mereka juga terpaksa menghadapi tekanan nilai sebagai pemeriksa yang dilabelkan sebagai pemurah atau kedekut. Di samping itu mereka juga terpaksa menjalankan kerja pemarkahan ke atas peruntukan wang kertas jawapan atau projek kerja kursus praktikal yang secara relatifnya adalah banyak dalam sesuatu masa. Sebagai manusia biasa, mereka memerlukan satu kekuatan tambahan untuk melaksanakan tugas murni ini dengan cara yang paling mungkin. Oleh yang demikian maka pengetahuan, pengalaman, bimbingan dari rakan-rakan pemeriksa yang berpengalaman dan maklumat-maklumat pemarkahan yang ada ini bagi melaksanakan aktiviti pemarkahan dengan berkesan adalah amat diperlukan.

Aktiviti pemarkahan kertas jawapan atau projek kerja kursus praktikal merupakan satu aktiviti yang diulang-

ulang oleh seseorang pemeriksa setiap tahun. Sepatutnya seseorang pemeriksa akan menjadi lebih cekap dan mahir apabila semakin kerap melakukan aktiviti pemarkahan. Berhubung dengan keadaan ini, kajian ini akan cuba untuk meninjau kecekapan pemeriksa dalam konteks mengulang aktiviti pemarkahan pada projek kerja kursus praktikal yang sama sebanyak dua kali ulangan. Adakah markah-markah diperolehi yang telah diberikan itu boleh mencapai tahap ketekalan dan kejituan yang diperlukan?

Kemungkinan sebenar seseorang pemeriksa juga terletak pada kecekapan dan ketelitian menggunakan skema pemarkahan. Ini adalah kerana skema pemarkahan merupakan satu kayu ukur yang membolehkan seseorang pemeriksa menghasilkan markah-markah diperolehi yang jitu jika skema itu digunakan dengan sebaik-baiknya. Oleh yang demikian jika beberapa orang pemeriksa yang berlainan yang bertugas dalam satu pasukan pemeriksa dapat menggunakan skema pemarkahan yang sama dengan berkesan maka jurang perbezaan markah diperolehi di antara mereka dapat dikurangkan ke tahap paling minimum. Sehubungan dengan itu maka kajian ini berharap dapat meninjau perbezaan markah-markah diperolehi yang mungkin berlaku di antara pemeriksa yang berlainan dalam keadaan di mana mereka memeriksa projek kerja kursus praktikal yang sama dengan menggunakan skema pemarkahan yang sama. Adakah kedudukan pemeriksa yang berlainan itu dapat menghasilkan markah-markah diperolehi yang tekal dan jitu?

Secara umumnya persepsi seseorang pemeriksa terhadap skema pemarkahan dan projek kerja kursus praktikal yang

periksa adalah berbeza-beza disebabkan oleh pelbagai faktor yang berkemungkinan boleh mempengaruhi mereka seperti taraf jawatan, pengalaman, pengetahuan, masa dan sifat skema itu sendiri sama ada maklumat-maklumat yang terkandung dalam skema pemarkahan itu secara terperinci (holistik) atau secara umum (global). Jika perbezaan persepsi ini dapat dikurangkan maka markah-markah diperolehi yang lebih jitu dan tekal dapat dihasilkan dengan baik sekali.

3 Pernyataan Masalah

Objektif utama kajian ini ialah untuk meninjau kejituan dan ketekalan markah-markah diperolehi yang telah diberikan pada projek kerja kursus praktikal MPKHPP pilihan B oleh pemeriksa yang bertugas sebagai guru MPKHPP pilihan B, pemeriksa yang bertugas sebagai penyetara MPKHPP pilihan B dan pemeriksa khas yang mempunyai latar belakang sebagai guru mata pelajaran Seni Perusahan. Mereka telah melaksanakan aktiviti pemarkahan pada projek kerja kursus praktikal yang sama dengan menggunakan skema dan kaedah pemarkahan yang sama. Oleh yang demikian maka kajian ini berharap dapat menjawab dua masalah berikut:

(1) Adakah terdapat sebarang perbezaan kejituan dan ketekalan pada markah-markah diperolehi bagi projek kerja kursus praktikal yang sama di peringkat pemarkahan intra-pemeriksa apabila mereka menggunakan skema pemarkahan yang sama? Adakah kemungkinan berlaku keadaan di mana pemeriksa yang sama memberikan markah diperolehi yang jauh

berbeza pada projek kerja kursus praktikal yang sama apabila mereka melaksanakan aktiviti pemarkahan pada masa yang berlainan? Jika ada perbezaan pada markah-markah diperolehi tersebut maka setakat manakah berlaku jurang perbezaan itu dan adakah markah-markah diperolehi yang diberikan itu jitu dan tekal?

(2) Adakah terdapat sebarang perbezaan kejituan dan ketekalan pada markah-markah diperolehi bagi projek kerja kursus praktikal yang sama di peringkat pemarkahan inter-pemeriksa apabila mereka menggunakan skema pemarkahan yang sama? Berhubung dengan soalan ini maka tiga sub-soalan berikut akan cuba dijawab oleh kajian ini:

(i) Adakah terdapat sebarang perbezaan kejituan dan ketekalan pada markah-markah diperolehi di antara pemeriksa dengan pemeriksa-penyetara?

(ii) Adakah terdapat sebarang perbezaan kejituan dan ketekalan pada markah-markah diperolehi di antara pemeriksa dengan pemeriksa khas?

(iii) Adakah terdapat sebarang perbezaan kejituan dan ketekalan pada markah-markah diperolehi di antara pemeriksa-penyetara dengan pemeriksa khas?

1.4 Hipotesis Kajian

Ada empat bentuk hipotesis nol yang telah dibina dan telah diuji pada aras signifikan $p < 0.05$. Hipotesis-hipotesis itu ialah:

(pemarkahan di peringkat inter-pemeriksaan) jika mereka menggunakan skema pemarkahan yang sama.

1.5 Definisi Istilah Operasional

1.5.1 Kebolehpercayaan Markah

Istilah kebolehpercayaan markah yang digunakan dalam kajian ini bermaksud ketekalan dan kejituan markah-markah diperolehi bagi projek bekas makanan haiwan. Ketekalan diperolehi dengan menggunakan rumus Cronbach's Coefficient Alpha dan rumus Pearson Product Moment Correlation Coefficient.

Rumus Cronbach's Coefficient Alpha yang telah digunakan ialah:

$$\alpha = \left(\frac{k}{k - 1} \right) \left(1 - \frac{\sum (SD_i)^2}{(SD_x)^2} \right)$$

- k - bilangan item-item dalam ujian
- SD_i - jumlah varians skor bagi semua pelajar pada satu item dan kemudian tambah semua varians daripada semua item-item
- SD_x - varians bagi jumlah skor ujian

Anggaran nilai kebolehpercayaan markah yang telah diperolehi adalah berdasarkan kepada satu keadaan ujian sahaja (a single form of the test). Rumus Cronbach's Coefficient Alpha telah digunakan dengan tujuan untuk meninjau 'ketekalan dalam' bagi markah-markah diperolehi iaitu markah bagi tiap-tiap item yang terdapat dalam skema pemarkahan. Adalah perlu dijelaskan bahawa prosedur

ini sensitif kepada sifat homogeneous item. Justeru itu nilai kebolehpercayaan yang diperolehi adalah tertakluk kepada faktor panjang ujian (length of the test) dan 'speed' (Nitko, 1983, ms. 396-397).

Seterusnya rumus Pearson Product Moment Correlation Coefficient yang telah digunakan ialah:

$$r = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N(\sum X^2) - (\sum X)^2][N(\sum Y^2) - (\sum Y)^2]}}$$

N - bilangan pelajar

X - skor (markah diperolehi)

Y - skor (markah diperolehi)

Rumus ini telah digunakan untuk meninjau darjah pertalian (korelasi) di antara dua pembolehubah (skor X dan skor Y) serta arah pembolehubah-pembolehubah itu berubah sama ada ke arah positif, negatif atau kosong (tiada pertalian di antara dua pembolehubah yang berkaitan). Faktor-faktor yang mungkin mempengaruhi nilai korelasi yang diperolehi ialah ralat pengukuran, had julat skor, darjah kepencongan taburan skor dan kurangnya pertalian linear di antara dua pembolehubah yang dikorelasikan. Nilai korelasi di antara pembolehubah X dan Y boleh meningkat tinggi sebab pembolehubah X mempengaruhi pembolehubah Y, atau pembolehubah Y mempengaruhi pembolehubah X atau pembolehubah Z mempengaruhi kedua-dua pembolehubah tersebut (Nitko, 1983, ms. 77-89).

Kejituan markah diperolehi pula telah dihitung dengan menggunakan kaedah analisis markah selisih iaitu dengan cara membandingkan di antara markah-markah oleh pemeriksa, pemeriksa-penyetara dengan markah-markah oleh pemeriksa khas.

1.5.2 Intra-Pemeriksa

Istilah intra-pemeriksa bermaksud korelasi bagi dua set markah-markah diperolehi yang telah diberikan oleh pemeriksa yang sama. Dua set markah diperolehi itu ialah markah MPS dan markah MPD. Nilai korelasi di antara markah MPS dengan markah MPD disebut sebagai nilai korelasi intra-pemeriksa.

1.5.3 Inter-Pemeriksa

Istilah inter-pemeriksa bermaksud korelasi markah-markah diperolehi di antara dua pemeriksa yang berlainan. Mereka telah memberikan markah pada projek-projek kerja kursus praktikal yang sama secara berasingan. Markah-markah diperolehi itu ialah markah MPS, MPP dan MPK. Nilai korelasi di antara markah MPS dengan markah MPP, markah MPS dengan markah MPK dan markah MPP dengan markah MPK disebut sebagai nilai korelasi inter-pemeriksa.

1.5.4 Pemeriksa

Pemeriksa ialah guru terlatih yang sedang mengajar PKHPP pilihan B. Di bawah konsep Penilaian Kendalian Sekolah Menengah (PKSM) mereka dikehendaki oleh Lembaga Pemeriksaan Malaysia dan Jabatan Pendidikan Kedah

menjalankan tugas sebagai pemeriksa kepada projek kerja kursus praktikal yang telah dibuat oleh pelajar mereka sendiri. Tugas sebagai pemeriksa ini dianggap sebagai sebahagian daripada tugas utama mereka. Untuk mencapai matlamat sebagai pemeriksa yang berkegangan maka mereka telah diberikan latihan pemarkahan yang mencukupi oleh Lembaga Peperiksaan Malaysia (Lembaga Peperiksaan, 1991).

1.5.5 Pemeriksa-Penyetara

Pemeriksa-penyetara ialah guru terlatih yang sedang mengajar MPKHPP pilihan B. Mereka telah dilantik untuk menjalankan tugas sebagai penyetara bagi projek kerja kursus praktikal MPKHPP pilihan B oleh Jabatan Pendidikan Kedah. Mereka juga telah menjalani kursus sebagai penyetara dari masa ke semasa. Kursus ini telah dianjurkan oleh Jabatan Pendidikan Kedah.

1.5.6 Pemeriksa Khas

Pemeriksa khas ialah guru terlatih yang sedang mengajar MPKHPP pilihan A (Seni Perusahaan). Beliau ialah seorang penyetara kepada projek kerja kursus praktikal bagi MPKHPP pilihan A. Dalam konteks kajian ini, penyelidik melantik beliau sebagai pemeriksa khas atas dasar pengalaman beliau sebagai guru, pemeriksa, penyetara dan penilai bagi projek kerja kursus praktikal mata pelajaran Seni Perusahaan. Kedudukan ini dianggap amat beresuaian dan secocok dengan pemeriksaan projek kerja kursus praktikal MPKHPP pilihan B khususnya projek bekas makanan haiwan.

1.5.7 Kerja Kursus Praktikal Mata Pelajaran Kemahiran Hidup Program Peralihan (MPKHPP) Pilihan B (Perdagangan & Keusahawanan)

Kerja kursus praktikal MPKHPP pilihan B ialah satu tugas yang wajib dibuat oleh pelajar yang telah memilih mata pelajaran ini bagi menduduki peperiksaan Sijil Rendah Pelajaran tahun 1992 (Pusat Perkembangan Kurikulum, 1989). Tugas bagi MPKHPP pilihan B yang telah dipilih oleh kajian ini ialah projek bekas makanan haiwan dari komponen bahagian kerja kayu dan logam.

1.6 Batasan kajian

Bagi mencapai tujuan kajian yang diperlukan maka had kajian ini adalah tertakluk kepada batasan-batasan berikut:

(1) Sampel projek yang telah dipilih ialah projek-projek bekas makanan haiwan yang telah dibuat oleh pelajar tingkatan tiga bagi MPKHPP pilihan B sahaja. Oleh hal yang demikian maka keputusan kajian yang telah diperolehi tidak dapat diitlakkan kepada semua projek kerja kursus MPKHPP pilihan B yang lain dan juga MPKHPP pilihan A, C dan D.

(2) Kajian ini menganalisis markah-markah diperolehi bagi projek bekas makanan haiwan sebagai markah diperolehi untuk produk akhir sahaja.

1.7 Andaian Kajian

(1) Pemeriksa, pemeriksa-penyetara dan pemeriksa khas ialah mereka yang telah mengikuti kursus pemarkahan

projek kerja kursus praktikal bahagian kemahiran manipulatif bagi MPKHPP.

(2) Pemeriksa, pemeriksa-penyetara dan pemeriksa khas telah melaksanakan aktiviti pemarkahan pada tahap sehabis baik yang mereka mampu menurut kebolehan dan kecekapan masing-masing.

(3) Projek kerja kursus MPKHPP pilihan B iaitu projek bekas makanan haiwan ialah projek yang asli, yang telah disediakan oleh pelajar yang terlibat di bengkel sekolah masing-masing dalam jangkamasa yang telah ditetapkan.

KAJIAN KARYA YANG BERKAITAN

2.1 Pengenalan

Bab ini memaparkan dapatan kajian dari karya-karya yang berkaitan dengan kajian kebolehpercayaan markah peperiksaan iaitu kebolehpercayaan intra-pemeriksa dan inter-pemeriksa. Seterusnya bab ini juga memaparkan kajian berhubung dengan unsur harapan guru, bentuk interaksi antara guru dengan pelajar bagi tujuan untuk meninjau pengaruh unsur tersebut terhadap markah yang diberikan dan juga unsur kecekapan guru sebagai pemeriksa. Secara umumnya, karya-karya yang berkaitan dengan kajian kebolehpercayaan markah kerja kursus praktikal didapati amat terhad bilangannya (Connoughton, 1969-1970; Lourdasamy & Siow Heng Loke, 1991) berbanding dengan kajian kebolehpercayaan markah peperiksaan bagi soalan yang memerlukan jawapan berbentuk esei. Oleh itu dapatan kajian yang telah dipaparkan dalam bab ini lebih banyak tertumpu kepada kajian kebolehpercayaan markah esei berbanding dengan kerja kursus bercorak praktikal. Sungguhpun demikian kajian-kajian tersebut (esei) telah dapat membantu penyelidik mendapatkan gambaran yang jelas berhubung dengan aspek kebolehpercayaan markah peperiksaan bagi tujuan mencapai objektif kajian.

2.2 Kebolehpercayaan Markah Peperiksaan

Willmott (1975) telah menyatakan bahawa kajian yang paling popular berhubung dengan kebolehpercayaan markah

peperiksaan ialah kajian oleh Hartog & Rhodes (1935). Hartog & Rhodes telah menjalankan kajian ke atas pelbagai mata pelajaran. Beliau mendapati bahawa ketidakbolehpercayaan markah peperiksaan bagi pemarkahan intra-pemeriksa dan inter-pemeriksa adalah berpunca dari kurang persetujuan di antara seorang pemeriksa dengan pemeriksa yang lain. Bagi pemeriksa yang sama pula ia adalah berpunca dari dua masa pemarkahan yang berlainan (Willmott, 1975). Kajian oleh Hartog & Rhodes ini kemudiannya telah diulangi oleh Pillinear & Connoughton (1969) dengan menggunakan sampel skrip esei bagi mata pelajaran English Composition. Pillinear & Connoughton telah mendapati bahawa nilai korelasi intra-pemeriksa ialah .90 dan inter-pemeriksa ialah .80. Nilai korelasi yang lebih rendah iaitu -.50 telah diperolehi apabila pemeriksa menggunakan 'kaedah pemarkahan impression' (dalam Willmott, 1975).

Seterusnya Christopher (dalam Willmott, 1975) telah menjalankan kajian bagi memperbaiki prestasi kebolehpercayaan markah peperiksaan dengan menggunakan 'kaedah penyetaraan markah'. Christopher telah merumuskan bahawa proses ini seharusnya dilakukan dengan cara mengumpulkan pemeriksa-pemeriksa sebaik sahaja peperiksaan tamat. Kemudiannya membincangkan skema pemarkahan secara terperinci, memeriksa dan membincangkan skrip sampel berpandukan kepada kriteria dan piawaian yang sama, menyetarakan markah-markah oleh pemeriksa yang kedekut atau yang pemurah, yang tidak tekal serta memeriksa semula skrip yang telah diberikan markah oleh pemeriksa untuk

menyemak keseragaman pemarkahan.

Burt (1945) telah mencadangkan pendeka' faktor' atau 'analisis varians' bagi mempe- kebolehpercayaan markah peperiksaan jika akti. pemarkahannya melibatkan penilaian yang bersifat subjektif. Beliau berpendapat bahawa kebaikan dari menggunakan pendekatan ini ialah markah-markah yang digunakan untuk memperolehi nilai kebolehpercayaan itu hanya melibatkan satu kali pemarkahan sahaja. Seterusnya beliau menyatakan bahawa adalah tidak praktikal serta sukar untuk mendapatkan pemeriksa yang rela mengulang aktiviti pemarkahan yang sama pada skrip jawapan dan calon yang sama jika menggunakan 'kaedah uji-ulang-uji'. Beliau telah menegaskan bahawa matlamat utama bagi meningkatkan tahap kebolehpercayaan markah peperiksaan ialah dengan cara membebaskan markah-markah dari ralat pengukuran. Ini adalah kerana salah satu punca ketidakbolehpercayaan markah ialah kandungan ralat pengukuran yang tinggi. Beliau telah mengemukakan tujuh cadangan bagi memperbaiki tahap kebolehpercayaan markah. Cadangan tersebut ialah:

(1) Butir-butir yang akan diberikan markah hendaklah dikenalpasti dengan betul dan didefinisikan dengan lengkap.

(2) Penilaian hendaklah berdasarkan kepada persembahan (performance) yang sebenar serta boleh diperhatikan dan bukan kepada impressi yang bersifat subjektif.

(3) Melibatkan sekurang-kurangnya

pemeriksa bagi memeriksa sekumpulan calon y'

h

(4) Memberi amaran kepada pemeriksa berhubung kesan halo.

(5) Markah hendaklah difikirkan sebagai satu ukuran yang relatif dan bukan sebagai satu ukuran yang mutlak.

(6) Pemeriksa tidak sepatutnya dibebankan dengan peruntukan terlalu banyak skrip jawapan dan terlalu banyak perbezaan butir-butir untuk diberikan markah bagi tujuan mengurangkan kesan hilang minat, letih serta kesuntukan masa.

(7) Apabila melibatkan ramai pemeriksa yang memeriksa skrip jawapan bagi calon yang sama maka tiap-tiap pemeriksa dicadangkan memeriksa secara bebas terlebih dulu sebelum berunding dengan lain-lain pemeriksa.

Wiseman (1949) seorang ketua pemeriksa telah menggunakan 'pendekatan uji-ulang-uji' untuk meninjau tahap kebolehpercayaan markah esei bagi kertas English Composition. Pemarkahannya menggunakan 'kaedah impression'. Beliau telah menjalankan kajian selama dua tahun yang melibatkan sampel sebanyak 2500 calon bagi tahun 1943 dan 1500 calon bagi tahun 1948. Skrip jawapan esei telah diperiksa oleh empat orang pemeriksa bebas. Dapatan kajian beliau telah menunjukkan bahawa nilai kebolehpercayaan markah bagi pemarkahan inter-pemeriksa ialah antara .72 hingga .85 bagi kajian pada tahun 1943 dan antara .60 hingga .86 bagi kajian pada tahun 1948.

Nilai kebolehpercayaan markah bagi pemarkahan intra-pemeriksa keseluruhannya ialah .80. Beliau telah mencadangkan supaya pemeriksa yang telah menunjukkan prestasi kebolehpercayaan kurang dari .70 digugurkan dari senarai pemeriksa. Seterusnya beliau telah mencadangkan supaya menggunakan pemarkahan inter-pemeriksa bagi tujuan mengawal keberkesanan pemeriksa.

Finlayson (1951) telah mengulangi kajian yang telah dijalankan oleh Wiseman (1949) dengan menggunakan 'pendekatan eksperimen'. Kajian beliau melibatkan sampel seramai 197 murid sekolah rendah. Tiap-tiap seorang murid dikehendaki menulis dua buah esei secara berasingan dalam dua jangkamasa yang berlainan. Esei ini kemudiannya diperiksa oleh 6 orang pemeriksa dalam keadaan mereka tidak mengetahui nama murid yang berkenaan. Nilai kebolehpercayaan markah esei yang telah diperolehi bagi intra-pemeriksa ialah antara -.80 hingga .60 dan inter-pemeriksa antara -.90 hingga .59. Dapatan kajian ini dapat menyokong kajian oleh Wiseman (1949). Beliau telah mencadangkan supaya menggunakan lebih ramai pemeriksa untuk meningkatkan kebolehpercayaan markah esei. Beliau juga telah menggunakan 'kaedah analisis varians' bagi melihat keberkesanan kaedah ini berbanding dengan 'kaedah uji-ulang-uji'. Nilai kebolehpercayaan markah yang telah diperolehi ialah .69 bagi intra-pemeriksa dan .86 bagi inter-pemeriksa. Beliau berpendapat bahawa tidak ada banyak perbezaan bagi nilai kebolehpercayaan markah yang ditunjukkan bagi kedua-dua kaedah yang digunakan itu.

Britton et. al. (1966) telah menjalankan kajian menggunakan 'pendekatan eksperimen' bagi meninjau kedudukan nilai kebolehpercayaan markah peperiksaan kertas English Composition peringkat GCE O-Level. Kajian mereka adalah bertujuan untuk melihat keberkesanan 'multiple marking' bagi kaedah pemarkahan analitik dan impression. Mereka telah menggunakan sampel sebanyak 500 skrip esei. Skrip esei ini telah diperiksa oleh 9 orang pemeriksa bebas. Nilai kebolehpercayaan markah esei yang telah diperolehi ialah intra-pemeriksa antara $-.55$ hingga $.25$ dan inter-pemeriksa antara $-.47$ hingga $.07$. Nilai ini adalah rendah dan tidak tekal dengan kajian oleh Wiseman (1949) dan Finlayson (1951). Mereka telah mendapati bahawa punca ketidakbolehpercayaan markah esei ialah kurang persetujuan antara seorang pemeriksa dengan pemeriksa yang lain. Pemeriksa secara individu telah menggunakan piawain sendiri secara tidak tekal. Adalah jelas menunjukkan bahawa kurang persetujuan antara pemeriksa dengan pemeriksa ketika melaksanakan aktiviti pemarkahan bagi skrip esei yang sama tetapi diperiksa oleh pemeriksa yang berlainan masih kekal sebagai satu faktor utama sejak 31 tahun yang lalu (kajian oleh Hartog & Rhodes, 1935) di mana faktor ini telah menyebabkan ketidakbolehpercayaan markah esei.

Backhouse (1972) dalam kajian yang bertajuk 'Reliability of GCE Examination: A Theoretical and Empirical Approach' telah mencadangkan tiga pendekatan bagi menganggarkan nilai kebolehpercayaan markah peperiksaan. Pendekatan-pendekatan tersebut ialah

'internal consistency reliability', 'analisis varians' 'mengkorelasikan markah bagi sesuatu soalan dengan markah ujian hipotetikal yang sama tara'. Beliau telah menjalankan kajian intensif ke atas 10 subjek di peringkat GCE A-Levels dan 9 subjek di peringkat GCE O-Levels termasuk kerja kursus praktikal. Tiap-tiap subjek telah diperiksa oleh seorang hingga tiga pemeriksa. 'Mark adjustment' oleh ketua pemeriksa dengan cara memeriksa semula skrip yang berkenaan telah digunakan untuk markah-markah yang menunjukkan kes 'borderline' serta 'scaling' untuk pemeriksa yang telah menunjukkan sikap kedekut atau pemurah memberikan markah. Dapatan dari kajian beliau telah menunjukkan nilai kebolehpercayaan markah bagi pemarkahan inter-pemeriksa ialah antara $-.06$ hingga $.88$ dan nilai yang rendah telah diperolehi bagi markah kerja kursus praktikal. Beliau telah merumuskan bahawa proses 'mark adjustment' dan 'scaling' yang telah dilakukan oleh ketua pemeriksa hanyalah penting dari segi keadilan tetapi tidak kepada kebolehpercayaan markah. Beliau nampaknya bersetuju dengan cadangan oleh Burt (1945) dan Stanley (1971) supaya menggunakan 'kaedah analisis varians' sebagai satu pendekatan yang sesuai bagi meninjau konsep kebolehpercayaan markah peperiksaan.

Willmott (1975) telah menggunakan kaedah 'internal consistency reliability' untuk meninjau kedudukan nilai kebolehpercayaan markah peperiksaan. Sampel kajian beliau terdiri daripada 83 keputusan peperiksaan CSE tahun 1969 dan 1970 serta 29 keputusan peperiksaan GCE tahun 1970 dan 1971. Dapatan kajian tersebut telah menunjukkan