

KESAHAN DAN KEBOLEHPERCAYAAN DALAM KAJIAN KUALITATIF DAN KUANTITA

DR. HJ. GHAZALI BIN HJ. DARUSALAM

Ketua Unit Penyelidikan,
Jabatan Penyelidikan dan Pembangunan
Maktab Perguruan Islam, Bangi

Pengenalan

Kesahan (validity) dan kebolehpercayaan (reliability) instrumen amat penting bagi mempertahankan kejituan instrumen daripada terdedah kepada kecacatan. Semakin tinggi nilai dan tahap kesahan dan kebolehpercayaan instrumen maka semakin jitu data-data yang akan diperoleh bagi menghasilkan kajian yang baik dan berkualiti. Kesahan dan kebolehpercayaan instrumen telah dipiawaikan oleh tiga agensi;

i. Committee of members of the American Educational Research

Association

- ii. The American Psychological Association dan
- iii. The National Council on Measurement in Education¹.

1. Kesahan Instrumen Soal Selidik (Kuantitatif).

Kesahan didefinisikan sebagai ketepatan (*appropriateness*), kebenaran (*truthfulness*), bermakna (*meaningfulness*) dan kebolegunaan (*usefulness*) instrumen yang membolehkan

¹ American Psychological Association (1985), *Standards for Education and Psychological Testing*. Washington D.C., : APA., h. 9-18 dan 19-23.

data-data diinferenkan². Kesahan juga bermaksud ; persetujuan antara dua percubaan untuk mengukur trait yang sama secara maksimum dengan kaedah yang berlainan³.

Biasanya bagi instrumen soal selidik, seseorang penyelidik akan menggunakan teknik *Criterion – Related Evidence* untuk mengumpulkan bukti (*collecting evidence*) terhadap kesahan instrumen melalui sekumpulan pakar rujuk (*expert judgment*)⁴. Dengan itu

² Fraenkel Jack. R dan Norman. E. Wallen, *op.cit.*, h. 153.

³ David Campbell dan David Fiske (1959), “ Convergent and Discriminant Validation”, *Psychological Bulletin*, Vol. 56, h. 81-105.

⁴ Fraenkel Jack. R dan Norman. E. Wallen, *op.cit.*, h. 154 - 159.

Terdapat tiga teknik dalam menentukan validity instrumen :

- i. Content-Related Evidence of validity
Ia merujuk kepada ketepatan kandungan dan format instrumen, sifat komprehensif instrumen, kebolegunaan variabel, ketepatan dan kecukupan kandungan item-item yang boleh memberi kefahaman kepada sampel untuk menjawabnya serta ketekalan kandungan dan format item untuk diukur dan dinilai oleh sampel.
Teknik ini menggunakan sekumpulan pakar rujuk (*expert judgment / intelligent judgment*) bagi membuktikan ketepatan kandungan item ketika menilai instrumen.
- ii. Criterion-Related Evidence of Validity
Ia merujuk kepada perhubungan antara skor yang diperolehi dari satu instrumen dengan skor yang diperolehi daripada satu atau lebih instrumen yang lain untuk melihat bagaimana kekuatan hubungan instrumen-instrumen tersebut.
Cara ini memerlukan penyelidik membuat ramalan (*predictive*) dan persetujuan (*concurrent*). *Predictive validity* bermaksud dua ujian perlu dibuat diawal dan pada akhir semester atau kursus kemudiannya dibandingkan skor antara kedua-duanya, jika kedua-
dua skor tidak jauh berbeza, maka ia dikatakan mempunyai *predictive validity*. *Concurrent validity* bermaksud dua ujian yang sama, ditadbir kepada dua sampel yang berbeza dalam serentak masa telah menghasilkan dua skor yang sama atau hampir sama. Teknik ini memerlukan bacaan *index correlation coefficient (r)*, yang menunjukkan darjah (*degree*) perhubungan yang wujud antara skor-skor yang diperolehi dalam dua instrumen.
- iii. Construct-Related Evidence of Validity
Ia merujuk kepada darjah pembuktian yang total yang diperolehi secara tekal daripada pelbagai bukti yang berbeza. Terdapat tiga langkah untuk mencari *Construct-Related Evidence* :
 - a. pembolehubah yang ingin diukur mestilah jelas maksudnya,
 - b. hipotesis yang dibina hendaklah berasaskan daripada teori yang disandarkan kepada pembolehubah,
 - c. hipotesis telah diuji secara logik dan empirik.
 Cara menentukan validitynya dengan melantik penilai bebas (*Independent Judges*) bagi menilai pembolehubah dan melihat kepada perhubungan nilai kolerasi antara dua skor.

penyelidik akan melantik beberapa orang pakar untuk mengesahkan instrumen kajian seperti jadual 1

Jadual 1 berikut menunjukkan senarai pakar rujuk bagi mengesahkan instrumen soal selidik.

Jadual 1

Kesahan Instrumen Soal Selidik Melalui Sekumpulan Pakar Rujuk

Bil	Pembolehkan Item	Panel Pakar Rujuk / Expert Judgment
1.	Matlamat kurikulum dan objektif Kursus Diploma Perguruan Malaysia, pengkhususan Pengajian Islam.	<ol style="list-style-type: none"> 1. Dr. Jusoh bin Kadir Jabatan Pengajian Agama Islam, Maktab Perguruan Islam, Bangi. 2. Dr. Kartini binti Baharum Penolong Pengarah Unit Kurikulum Bahagian Pendidikan Guru, Kementerian Pendidikan Malaysia.
2.	Strategi dan kaedah pengajaran dan pembelajaran Kursus Diploma Perguruan Malaysia, pengkhususan Pengajian Islam.	<ol style="list-style-type: none"> 1. Dr. Hamzah bin Ihsan Penolong Pengarah Unit Dasar, Perancangan dan Pembangunan Bahagian Pendidikan Guru, Kementerian Pendidikan Malaysia 2. Dr. Jusoh bin Kadir Jabatan Pengajian Agama Islam, Maktab Perguruan Islam, Bangi.
3.	Produk dan output kursus Diploma Perguruan Malaysia, pengkhususan Pengajian Islam	<ol style="list-style-type: none"> 1. Dr. Kartini bini Baharum Penolong Pengarah Unit Kurikulum, Bahagian Pendidikan Guru, 2. Dr. Jusoh bin Kadir Jabatan Pengajian Agama Islam, Maktab Perguruan Islam, Bangi.
4.	Permasalahan Kursus Diploma Perguruan Malaysia, pengkhususan Pengajian Islam.	<ol style="list-style-type: none"> 1. Dr. Faridah binti Mohamed Penolong Pengarah Hal Ehwal Guru Pelatih, Bahagian Pendidikan Guru, Kementerian Pendidikan Malaysia.

2. Dr. Jusoh bin Kadir
Jabatan Pengajian Agama Islam,
Maktab Perguruan Islam,
Bangi.

2. Kesahan Instrumen Temu Bual Dan Pemerhatian (Kualitatif)

Bagi kajian penyelidikan yang menggunakan pendekatan kualitatif melalui temu bual, pemerhatian dan dokumen dalam usaha pengkaji untuk mengumpulkan sebanyak mungkin maklumat atau data, penggunaan teknik triangulasi itu sendiri telah dapat memperkukuhkan kesahan kajian⁵. Penggunaan teknik triangulasi (soal selidik, temu bual dan pemerhatian) yang digunakan secara serentak dalam sesuatu kajian hendaklah dibuat dalam bentuk *equivalent form* (kriteria dan konstruk item yang sama dalam pelbagai jenis instrumen).

Penggunaan teknik triangulasi seperti pungutan data-data yang diperoleh melalui sumber dokumen bukan sahaja boleh menyokong dan menambahkan bukti daripada pelbagai sumber lain, tetapi juga menolong mengesahkan maklumat yang didapati daripada temu bual dan pemerhatian⁶.

Kesahan dalam kajian berbentuk kualitatif (penggunaan temu bual, pemerhatian dan dokumen sebagai instrumentasi) tidak hanya tertumpu kepada hasil kajian tetapi keseluruhan

⁵ John Van Maanen (1983), *Qualitative Methodology*. Beverly Hill, California : Sage Publication, h. 136.

⁶ Suseela Malakolunthu, “Pengumpulan dan Analisis Data Kualitatif : Satu Imbasan”, dalam Marohaini Yusoff (2001), *Penyelidikan Kualitatif : Pengalaman Kerja Lapangan Kajian*. Kuala Lumpur : Universiti Malaya, h. 131.

proses penyelidikan⁷. Miles dan Huberman (1994) berpendapat tidak ada peraturan tertentu dalam pembinaan keesahan dalam kajian kualitatif⁸.

Tetapi McCall dan Simmons (1969) menegaskan, kajian kualitatif melalui pungutan data-data daripada temu bual dan pemerhatian juga memerlukan keesahan data⁹.

Campbell (1955) mencadangkan untuk menyelesaikan masalah kesahan pada data-data kualitatif khususnya temu bual ialah dengan menggunakan teknik *morale ballot*¹⁰. Manakala Bailey (1978) dan Scoot (1961) mencadangkan penggunaan teknik *independent rating* melalui *cross-checking* untuk membuat keesahan data kualitatif¹¹.

⁷ Steinar Kvale (1997), *The Social Construction of Validity*. Kerja-Kerja European Research of Teaching and Learning, Greece, h. 6-7, dalam Marohaini Yusoff (2001), *op.cit.*, h. 336.

⁸ Matthew B. Miles and Michael A. Huberman, (1994), *Qualitative Data Analysis : A Sourcebook of New Methods*. Beverly Hills : Sage Publication, h. 134.

⁹ George J. McCall dan J.L. Simmons (1969), *Issues in Participant Observation : A Text and Reader*. Reading Mass : Addison-Wesley Pub. Co., h. 19-23.

¹⁰ David Campbell (1955), "The Informant in Qualitative Research", *American Journal of Sociology*, vol, LXI, Nos 1-6, July 1955 – May 1956, Chicago, Illinois : University of Chicago Press, h. 339-342.

Morale Ballot bermaksud; menggunakan informant (pemberitahu) sebagai alat kajian untuk mendapatkan validiti data yang dibuat secara membuat rating dan membandingkan dua garisan informan melalui beberapa pengertian yang sama maknanya.

¹¹ Kenneth D. Bailey (1978), *Methods of Social Research*, New York : Free Press, h. 155.

Juga dalam Chirstopher Scoot (1961), "Research on Mail Surveys", *Journal of the Royal Statistical Society*, vol. 124, Series A (General), London : Royal Statistical Society, h. 186.

Mereka berpendapat validiti terhadap data temu bual dan pemerhatian boleh dilakukan dengan membuat independent rating secara cross-checking terhadap maklumat yang dipungut daripada informan pertama kepada informan kedua, sama ada dalam baris yang sama atau baris berlainan terhadap maklumat yang sama dalam teknik pungutan data yang sama (temu bual dengan temu bual). Jika maklumat yang diperolehi sama antara kedua-dua informan, maka teknik ini boleh membantu meningkatkan darjah validiti. Contohnya cross-checking ketepatan maklumat antara x13 dengan x23

abb

Informan pertama : baris 1 x11 -----x12-----x13-----x14-----x15

abb

Informan kedua : baris 2 x21-----x22-----x23-----x24-----x25

Bagi seseorang penyelidik yang menggunakan kaedah survey, maka kesahan dalaman (*internal validity*) dan kesahan kriteria (*criteria validity*) digunakan serentak bagi maksud membuat kesalahan kepada instrumen temu bual. Kesahan dalaman dilihat daripada aspek manipulatif kawalan dalam hubungan variabel-variabelnya. Dengan itu, teknik *morale ballot* dan *cross-checking* digunakan kedua-duanya serentak bagi memperkukuhkan nilai kesahan terhadap data, khususnya dalam temu bual¹². Atas saranan ini, biasanya penyelidik akan menggunakan dua orang informan (seorang lelaki dan seorang wanita) bagi setiap kajian bagi tujuan temu bual.

Manakala kesahan kriteria dalam temu bual dicapai melalui menemu bual individu yang benar-benar memiliki pengetahuan dan maklumat terhadap bidang kajian pengkaji¹³. Biasanya bagi tujuan ini pula, penyelidik akan menemu bual beberapa orang informan dalam lokasi kajian yang berlainan asalkan sama ciri dan sifatnya.

Manakala kesahan dalam pemerhatian dibuat melalui teknik kesahan konstruk (*construct validity*) iaitu dengan cara mengkategorikan peristiwa-peristiwa yang diperhatikan kepada beberapa kelompok (*clusters*) bagi variabel yang dikehendaki, kemudian membandingkan pemerhatian yang lalu dengan pemerhatian sekarang. Untuk meninggikan

¹² Merton, K. dan Paul Kandall (1946), "The Focussed Interview", *American Journal of Sociology*, vol. 51., Nos 1-6., July 1945 – May 1946, Chicago, Illinois : University of Chicago Press, h. 544.

Beliau menjelaskan, temu bual yang berkesan dan berjaya dibuat dengan dua cara :

- i. Temu bual dengan mendapatkan data-data yang mengandungi unsur-unsur validity dan reliability.
- ii. Temu bual dibuat dengan amat memuaskan dan telah memenuhi rasa ingin tahu penyelidik.

¹³ James Coleman (1958), "Relational Analysis : The Study of Social Organization with Survey Method", *Human Organization*, vol. 17., h. 28-36.

darjah kesahan konstruk, pemerhatian perlu dibuat dengan melihat fenomena yang sama dengan maksimum perbezaan dimensi atau perspektif¹⁴.

Penyelidik biasanya akan membuat pemerhatian kepada beberapa orang informan dengan mengkategorikannya mengikut fokus kajian. Kemudiannya membuat kelompok terhadap isi teks (transkripsi) yang digunakan. Manakala untuk meninggikan darjah kesahan konstruk, penyelidik akan membuat pemerhatian sebanyak lima hingga enam kali bagi setiap seorang informan selama beberapa hari. Data-data yang diperoleh melalui pemerhatian (bagi pengajaran guru di kelas) akan diproses dengan menggunakan teknik segmen-segmen pengajaran.

Kesahan criteria boleh diperoleh dalam pemerhati turut serta (participant observation), penyelidik perlu membuat semakan terhadap maklumat-maklumat yang dikumpulkan sama ada data tersebut penuh atau data tidak penuh¹⁵. Oleh sebab proses pemerhatian pengajaran informan dibuat sebanyak lima hingga enam kali dengan selang masa dan hari yang berlainan, maka seseorang penyelidik bolehlah berpendapat bahawa data-data yang dikehendaki telah mencukupi.

Tujuan penyelidik membuat kesahan instrumen adalah bagi menjamin keesahan item-item (instrumen) agar bersifat ;

- i. Defensibility (dapat dipertahankan – hasil kajiannya tepat dan berguna)
- ii. Accuracy (ketepatan – menjawab persoalan kajian)
- iii. Appropriateness (kesesuaian – relevan dengan tujuan kajian)
- iv. Meaningfulness (bererti – memberi makna terhadap data melalui skor) dan

¹⁴ Bailey K.D., (1978), *op.cit.*, h. 155.

¹⁵ Christopher Scott (1961), *op.cit.*, h. 168.

- v. Usefulness (kebolehgunaan – mampu membuat keputusan berhubung dengan apa yang cuba dicari atau dihasilkan kerana keputusan dari penilaian tersebut boleh menyediakan maklumat yang bermakna mengenai tajuk dan pembolehubah yang hendak diukur untuk menginferenkan data kajian)¹⁶.

3 Kebolehpercayaan Instrumen Soal Selidik (Kuantitatif)

Kebolehpercayaan instrumen ialah suatu nilai ukuran untuk menentukan ketekalan skor setiap item¹⁷. Ketekalan bermaksud apabila item yang sama diuji beberapa kali kepada subjek yang sama pada selang masa yang berlainan tetapi tetap memberi skor keputusan atau jawapan yang sama atau hampir sama¹⁸. Mengikut Campbell dan Fiske (1959), reliabiliti diertikan sebagai; persetujuan antara dua usaha untuk mengukur sifat (trait) yang sama secara maksimum melalui kaedah yang sama¹⁹.

Konsep pengukuran reliabiliti dalam kaedah kuantitatif khasnya penggunaan soal selidik direka untuk menguji kelompok soal selidik melalui kajian rintis (*pilot test*) yang bererti; satu ujian skala kecil (*small-scale testing*)²⁰. Kajian rintis juga berupa permulaan percubaan (*preliminary trial*) sebelum item-item daripada ujian sebenar dikenakan pada sampel sebenar. Tujuan kajian rintis dibuat adalah untuk mendapatkan ketelusan data

¹⁶ Fraenkel Jack. R dan Norman. E. Wallen, *op.cit.*, h. 153.

¹⁷ William Wiersma (2000), *op.cit.*, h. 297.

Beliau menjelaskan ; *Reliability means consistency, an observed score can be seen as consisting of two parts, one part in individual's "true" score and the other part an " error " score.*

¹⁸ Howard Wainer & Henry I. Braun (1988), *Test Validity*. New Jersey : Lawrence Erlbaum Associates Publishers, h. 20-25.

¹⁹ David Campbell dan David Fiske (1959), *op.cit.*, h. 81-105.

²⁰ David Pratt (1980), *op.cit.*, h. 414.

daripada ujian percubaan melalui kumpulan kecil individu²¹. Tujuan lain adalah untuk menilai ketekalan (*reliability*) item daripada aspek; aras item, objektif item, kefahaman item, kebolegunaan item dan arahan item itu sendiri²².

Jumlah sampel atau responden yang digunakan dalam kajian rintis biasanya tidak kurang daripada 20 orang sampel²³. Instrumen yang baik adalah instrumen yang dibina sendiri oleh penyelidik bagi maksud pengumpulan data dalam sesuatu kajian. Dengan itu, seseorang penyelidik tidak begitu digalakan untuk mengambil sebarang contoh, bentuk dan jenis soal selidik yang sedia ada atau yang dikeluarkan oleh agensi atau bank item daripada *sources of test information* bagi penyelidikan pendidikan atau daripada pihak pengeluar dan pembekal soal selidik seperti; Mental Measurements Year books (MMY)²⁴.

Penyelidik juga tidak digalakkan mengambil daripada agensi yang membekalkan ujian dalam pelbagai bidang penyelidikan seperti; Ujian Personaliti, Kecergasan, Tingkah Laku, Keberkesanan, Pendidikan, Kaunseling, Psikologi, Perniagaan, Pentadbiran dan

²¹ Walter R. Borg & Meredith Damien Gall (1979), *Educational Research ; An Introduction*. 3rd Edition, New York : Longman, h. 70

²² Gale H. Roid & Thomas M. Haladyna (1982), *A Technology for Test-Item Writing*. New York : Harcourt Brace Jovonovich Publishers, h. 206.

²³ Robert J. Mc Dermoot & Paul D. Sarrela (1996), *Health Education Evaluation and Measurement : A Practitioner's Perspective*. 2nd Edition, New Jersey : Prentice Hall Englewood, h. 147-148. Dermoot dan Sarrela pernah menggunakan 20 sampel dalam pilot test, melalui method test-retest. Beliau mentafsirkan indek kebolehpercayaan sebagai ; ... *the difficulty index estimates the difficulty of an item on an achievement test (e.g an item might have a difficulty index of a $\alpha = .80$, indicating that 80 percent of the subject answered the item correctly).*

²⁴ Conoley J.C, dan Kramer J.J. (1989), "The Tenth Mental Measurements Yearbook" . Lincoln, NE : University of Nebraska Press dalam L.R. Gay dan Diehl P.L (1992), *Research Methods for Business and Management*. New York : Mac Millan Publishing, Company, h. 170.

sebagainya yang dikeluarkan oleh PRO-ED Publications yang menjual lebih daripada 3,000 ujian dalam pelbagai bidang²⁵.

Alasan mengapa seseorang penyelidik tidak begitu digalakkan membeli soal selidik yang dikatakan standard dalam sesuatu kajian kerana terbukti sebagai '*well-business of selling test*' dengan meletakkan nilai pekali reliability coefficient $\alpha = > .90$ sebagai iklan, sedangkan ia sebenarnya tidak boleh dipertanggungjawabkan²⁶.

Alasan lain ialah kerana item-item berkenaan hanya sesuai digunakan pada masa tersebut sahaja dan menjadi tidak relevan pada masa kajian kini yang sedang dijalankan di samping kriteria-kriteria dalam *item construct (total test)* dan indikator item (sub item) yang berbeza²⁷.

Dengan itu, penyelidik sepatutnya menguji sendiri kebolehpercayaan (reliability) instrumen soal selidik melalui kajian rintis dengan menggunakan salah satu teknik atau

²⁵ Daniel J. Kayser dan Richard C. Sweetland (1985), *Test Critiques*, Vol. III. Kansas City : Test Corporation of America, h. 164.

Juga boleh di dapati dalam Daniel J. Kayser dan Richard C. Sweetland (1990), "Test : A Comprehensive Reference for Assessment in Psychology", Education and Business, 3rd ed., Austin, TX : PRO-ED dalam L.R. Gay dan Diehl P.L (1992), *Research Methods for Business and Management*, New York : Mac Millan Publishing, Company, h. 183.

²⁶ L.R Gay (1992), *op.cit.*, h. 169.

seperti; Self-Completion Questionnaire (SCQ), Perception of Success Questionnaire (PSQ), Standard Teacher Efficacy Questionnaire (STEQ), Self-Administered Questionnaire (SAQ) dalam bentuk inventory, Questionnaire –On-Teacher-Interaction (QTI), Strengths-and-Difficulties-Questionnaire (SDQ), Communication Climate Measure (CCM), Study Group Evaluation Questionnaire (SGEQ), Individualized dan Classroom Environment Questionnaire (ICEQ)

²⁷ Steven J. Osterlind (1992), *Constructing Test Item*. 2nd Edition, London : Kluwer Academic Publishers, h. 46.

kombinasi antara beberapa teknik untuk menguji darjah kebolehpercayaan item yang bertujuan untuk melihat *internal consistency correlation coefficient* item-item²⁸.

²⁸ Gale H. Roid & Thomas M. Haladyna (1982), *A Technology for Test-Item Writing*. New York : Harcourt Brace Jovonovich Publishers, h. 206.

Lihat juga dalam L.R Gay (1992), *Educational Research : Competencies for Analysis and Application*. 4th Edition, Maxwell Macmillan International Edition, New York : Maxwell Macmillan Publishing Company, h. 168-169.

Lihat juga dalam Sheridan J. Coakes & Lyndall G. Steed (2001), *SPSS Analysis Without Anguish*. Brisbane : John Willey & Sons Ltd., h. 147.

Lihat juga dalam Mohd. Salleh Abu dan Zaidatun Tasir (2001), *Pengenalan Kepada Analisis Data Berkomputer SPSS 10.0 For Windows*. Kuala Lumpur : Venton Publishing. h. 252-255

Pada umumnya terdapat tiga cara untuk menguji darjah kebolehpercayaan item (test):

i. Test-Retest Reliability.

Ia menjelaskan darjah skor yang sentiasa konsisten setiap masa apabila ia diuji. Instrumen yang sama dinilai secara berulang kali kepada kumpulan sampel yang sama, pada masa yang berlainan. Dua set data yang diperoleh akan dianalisis menerusi analisis korelasi untuk melihat kekuatan perhubungannya. Sekiranya pekali kolerasi (coefficients reliability) menghampiri nilai 1, maka kedua-dua set data tersebut mempunyai perkaitan atau perhubungan yang kuat (instrumen mempunyai nilai kebolehpercayaan yang tinggi).

ii. Equivalent-Form Reliability

Dua set test yang berbeza tetapi mempunyai kandungan yang sama (variabel), sama jumlah item, sama struktur, sama aras kesukaran dan sama arahan). Kedua-dua instrumen tersebut akan diberikan kepada kumpulan sampel yang sama pada masa yang sama. Dua set data yang diperolehi akan dianalisis cara korelasi untuk mendapatkan pekali kolerasi (reliability coefficients). Skor yang diperolehi daripada kedua-dua sampel hendaklah dikoreletkan. Purata skor yang diperolehi daripada kedua-dua test ialah petunjuk kepada darjah reliabiliti.

Ia juga disebut sebagai Alternate-Form Reliability yang menunjukkan kepelbagaian skor dari satu form kepada satu form. Test-Retest dan Equivalent-Form Reliability boleh dikombinasikan bagi menilai stabiliti skor yang diperolehi pada setiap masa dan ditadbirkan kepada sejumlah sampel kecil. Teknik ini amat baik kerana item-item boleh digeneralisasikan dalam satu set item.

iii. Internal Consistency

Kaedah ini digunakan bagi menilai ketekalan dalaman instrumen. Hanya satu set item sahaja yang diuji kepada sejumlah kecil sampel. Terdapat tiga kaedah :

- a. prosedur Split-Half : bagi instrumen yang panjang dan sukar ditadbir. Ia memerlukan pembahagian kepada dua kategori mengikut nombor genap dan nombor ganjil item, ditadbir pada dua masa yang berlainan dengan dua form test pada satu kumpulan sampel. Rumus Spearman-Brown Correction digunakan.
- b. prosedur Kuder-Richardson : Ia juga disebut sebagai; Rationale Equivalent Reliability bagi menentukan bagaimana kesemua item dalam satu test berhubungan antara satu item dengan item yang lain, antara sub item dengan total item. Formula KR-20 atau KR-21 akan digunakan bagi menilai reliabilitinya.
- c. prosedur Cronbach Alpha : Ia lumrah digunakan bagi menilai pekali kolerasi koefisien yang berasaskan kepada purata korelasi item-item dalam satu test jika item-item tersebut berbentuk standard. Jika Item-item tidak standardised, maka purata covarian akan digunakan dalam analisisnya. Standardised item Alpha adalah dibuat bagi memperbandingkan varian perbezaan di antara dua Alpha, Alpha dan Standardised Item Alpha.

Pentafsiran pekali kebolehpercayaan yang boleh diterima mengikut pengamal penyelidikan dalam sains sosial ialah lebih daripada $\alpha = .60$. Dengan menggugurkan butiran-butiran yang merendahkan pekali tersebut (dalam construct dan indicator item), penyelidik boleh meningkatkan nilai reliabiliti terhadap alat ukur. Mengikut Mohd. Majid Konting (1993) tahap $\alpha = 0.71 - 0.99$ adalah tahap yang terbaik (71% - 99% kebolehpercayaan item oleh sampel)²⁹. Fraenkel dan Wallen (1996) meletakkan nilai reliability item yang diterima pada tahap $\alpha = 0.70 - 0.99$ ³⁰. Kubiszyn dan Borich (2000) menentukan nilai $\alpha = .80 - .90$ *reliability coefficients* yang diterima³¹. Manakala Popham (1990) pula ialah pada tahap $\alpha = .90 - .95$ ³². Tafsiran nilai cronbach alpha (α) *correlation coefficients* adalah seperti berikut:

0.00	hingga	+ 1.00	= pada asasnya ³³
.60	hingga	.70	= satisfied coefficients
.70	hingga	.80	= stability coefficients
.80	hingga	.90	= customary coefficients
.90	hingga	.95	= sufficient coefficients ³⁴
.80	hingga	.90	= acceptable reliability
.90	hingga	+ 1.00	= very good reliability ³⁵
.95	hingga	+ 1.00	= acceptable standardised test for internal consistency ³⁶ .

²⁹ Mohd. Majid Konting (1993), *op.cit.*, h. 182.

³⁰ Fraenkel J.R dan Norman E. Wallen (1996), *op.cit.*, h. 163.

³¹ Tom Kubiszyn & Gary Borich (2000), *Educational Testing and Measurement : Classroom Application and Practice*. Sixth Edition. New York : John Wiley & Sons, Inc., h. 319.

³² James Popham (1990), *Modern Educational Measurement. A Practitioner's Perspective*. 2nd Edition, New Jersey: Prentice Hall, Englewood Cliffs, h. 127.

³³ Howard B. Lyman (1986), *Test Scores and What They Mean*. Fourth Edition, New Jersey : Prentice Hall Englewood Cliffs, h. 29. Juga boleh dilihat dalam, William Wiersma (2000), *Research in Education : An Introduction*. Boston : Allyn and Bacon, h. 297.

Reliability coefficients can take one values from -0.1 to +1.0 inclusive conceptually, if a reliability coefficient were 0, there would be no 'true' component in the observed score. The observe score would consist entirely of error. On the other hand, if the reliability coefficient were 1.0 the observed sore would contain no error. 1.0 consist entirely of the true score, although coefficients of 1.0 are very rare inded

³⁴ Walter R Borg & Meredith Damien Gall (1979), *Educational Research : An Introduction*. 3rd Edition, New York, London : Longman, h. 97.

³⁵ Vierra, A. & Pollock, J. (1992), *Reading Education Research*. Scottsdale : Gorsuch Sewishirch, h. 86.

³⁶ Tom Kubiszyn & Gary Borich (1993), *Educational Testing and Measurement*. New York : Harper Collins, h. 166.

Pada amalnya kebanyakan penyelidik menggunakan Cronbach's Coefficient Alpha (α) untuk mengukur kebolehpercayaan item-item soal selidik (menggunakan skala Likert 5). Cronbach's Alpha sesuai digunakan bagi :

- i. menentukan ketekalan dalaman sesuatu konstruk (pembolehubah utama) item dan sesuatu indikator (pembolehubah kecil) dalam sub item,
- ii. item-item yang berbentuk aneka pilihan jawapan yang diberi nilai 3 atau lebih poin (bukan untuk dichotomous item),
- iii. mengira pekali kebolehpercayaan bagi data yang dipungut secara single occasion atau pungen data secara cross-sectional dalam kajian survey.

Nilai pekali kebolehpercayaan instrumen sangat bergantung kepada, homogeniti item dan isi kandungan item. Formula atau rumus statistik yang digunakan ialah:

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

- k : *bilangan item*
 $(SD_i)^2$: *variance of total scores (variance of the scores of all the samples on one item)*
 $\sum (SD^2)^2$: *sum of item variances*³⁷

Sebagai contohnya laporan hasil dapatan daripada kajian rintis 1 dan 2 adalah seperti jadual 2 berikut :

³⁷ Amir bin Mohd. Salleh, Ketua Penolong Pengarah, Unit Penyelidikan, Bahagian Perancangan dan Penyelidikan Dasar Pendidikan (EPRD), Kementerian Pelajaran Malaysia. Penyelidik belajar dengan beliau di biliknya pada 8. 12. 2002.

Jadual 2
Keputusan Kajian Rintis 1 dan Kajian Rintis 2

No Item	Item	Variabel	Nilai Alfa (α)	
			Ujian 1	Ujian 2
1 – 6	Demografi	var 1 – var 17	-	-
7 – 18	<u>Dimensi context</u> : Matlamat kurikulum dan objektif kursus DPM, PA	var 18 – var 45	$\alpha = .9642$	$\alpha = .9683$
19 - 20	<u>Dimensi input</u> : Strategi dan kaedah pengajaran	var 46 – var 60	$\alpha = .8545$	$\alpha = .9211$
21 - 29	<u>Dimensi product</u> : Output kursus	var 61 – var 69	$\alpha = .8804$	$\alpha = .9737$
30 – 40	<u>Dimensi proses</u> : Permasalahan kursus	var 70 – var 80	$\alpha = .7549$	$\alpha = .8533$

Hasil keputusan Ujian 1 dalam kajian ini, menunjukkan indeks nilai kebolehpercayaan berada di antara *cronbach $\alpha = .75$ (stability coefficients)* hingga *.96 (very good reliability atau acceptable standardised test for internal consistency)* masih berada pada tahap *degrees of positive correlation* (seperti yang ditunjukkan dalam *scatterplots* Fraenkel). Darjah pekali *correlation coefficient cronbach $\alpha = .75$* (soalan kajian ke empat : permasalahan kursus) menurun, disebabkan dalam satu set soal selidik terdapat banyak *total test* (item utama) di samping mengandungi *sub test* (item sub)³⁸.

L.R Gay (1992) menjelaskan jika satu set soal selidik mempunyai sub test (dalam jumlah yang agak banyak), maka nilai r (correlation) tidak akan mencapai tahap Cronbach α

³⁸ Robert E. Stake (1968), "Toward a Technology for the Evaluation of Educational Program" dalam. Ralph W. Tyler, Robert M. Cagne & Michael Scriven (1968), *Perspectives of Curriculum Evaluation*. Chicago : Rand Mc Nally & Company. h. 6.
Mengikut Stake, indeks kebolehan item boleh jatuh jika sub item terlalu banyak dalam sesuatu item.

= .90, dengan itu nilai Cronbach $\alpha = .65$ dikira pada tahap *satisfied coefficients*³⁹. Manakala Conrley & Impara (1995) menjelaskan, standard nilai item yang mengandungi banyak *sub test* biasanya berada di antara Cronbach $\alpha = .61$ hingga .92, manakala item-item dalam *total test* biasanya berada di antara Cronbach $\alpha = .85$ hingga .90 sahaja⁴⁰.

Biarpun begitu, analisis kajian rintis 2 yang dibuat tiga minggu kemudiannya (setelah item-item dalam *pool alpha if item deleted* dalam kajian rintis 1 sama ada dibuang atau diperbaiki) telah menunjukkan peningkatan indeks kebolehpercayaan yang tinggi cronbach $\alpha = .85$ hingga .97 yang berada pada tahap *acceptable* dan *very good reliability*. Dengan itu seseorang penyelidik tidak perlu lagi mengulang kajian rintis setelah merasa berpuas hati terhadap kebolehpercayaan instrumen sebelum instrumen ditadbir kepada sampel sebenar.

Kesahan dan kebolehpercayaan instrumen amat penting bagi mempertahankan kejituan instrumen daripada terdedah kepada kecacatan. Semakin tinggi nilai dan tahap kesahan dan kebolehpercayaan instrumen maka semakin jitu data-data yang akan diperolehi bagi menghasilkan kajian yang lebih baik dan berkualiti.

Seseorang penyelidik merujuk kepada beberapa orang panel penilai (pakar) untuk menyemak reka bentuk soal selidik dan indeks reliabiliti sesuatu instrumen.

³⁹ Walter R. Borg & Meredith Damien Gall (1979), *op.cit.*, 97.

⁴⁰ Jane Clase Conrley & James C. Impara (1995), *Thirteenth Mental Measurements Year Book*. Lincoln : University of Nebraska Press, h. 299.

membuat pemerhatian langsung maka ia boleh menambah nilai kebolehpercayaan maklumat-maklumat pemerhatian⁴³. Sebagai contohnya seseorang penyelidik boleh menggunakan teknik pemerhatian langsung ini melalui kajian lapangan terhadap lapan orang guru pelatih KDPM, PA (empat lelaki dan empat perempuan) di lapan buah sekolah yang berlainan kawasan dan negeri :

- i. Sekolah Kebangsaan Jalan 4, Bandar Baru Bangi, Selangor.
- ii. Sekolah Kebangsaan Seri Sekamat, Jalan Cheras, Kajang, Selangor.
- iii. Sekolah Kebangsaan Kampung Bakam, Miri, Sarawak.
- iv. Sekolah Kebangsaan Agama Bandar Miri, Miri, Sarawak.
- v. Sekolah Kebangsaan Kampung Apin-Apin, Keningau, Sabah.
- vi. Sekolah Kebangsaan Bunga Raya, Keningau, Sabah.
- vii. Sekolah Kebangsaan Kinasabaran, Kota Belud, Sabah.
- viii. Sekolah Kebangsaan Pekan Tuaran, Tuaran, Sabah.

Satu lagi cara untuk mengukur kebolehpercayaan pemerhatian ialah melalui dua teknik berikut:

- i. *Tandem interview* iaitu pengulangan pemerhatian (*re-observation*) dengan selang masa yang singkat (1 minggu) terhadap informan yang sama iaitu pemerhatian dibuat secara pertentangan peristiwa dengan item yang sama. Teknik ini telah juga digunakan terhadap lapan orang informan KDPM, PA
- ii. Melakukan kedua-dua jenis pemerhatian; pemerhatian berstruktur (variabel ditentukan terlebih dahulu) dan pemerhatian turut serta (*participant observation*)⁴⁴.

⁴³ Marlene De Laine (2000), *Fieldwork, Participation and Practice : Ethics and Dilemmas in Qualitative Research*. London : Thousand Oaks, California : Sage, h. 112.

⁴⁴ ¹⁵¹ Mak Lau Fong (1981), *Evaluating Validity and Reliability of Selected Qualitative Research Methods*. Singapore : National University of Singapore. h. 12

Tandem bererti sebuah basikal yang dikayuh oleh dua orang.

Lihat dalam, Robert Vales (1950), *Intercarion Process Analysis*. Reading, Ma : Addison-Wesley, h. 47

Juga dalam, Eugene J. Webb, Campbell, D.T., Schwartz, R.D and L. Sechrest (1966), *Unobtrusive Measures*. Chicago : Rand Mc. Nally, h. 86.

Terdapat 6 cara pemerhatian yang dicadangkan:

- i. Structured observation : jika informan diberi sesuatu dan ditempatkan di sebuah bilik, penyelidik pula menggunakan senarai semak item.
- ii. Unstructured observation : memerhatikan informan di satu tempat
- iii. Semi-structured observation : jika penyelidik menggunakan senarai semak untuk membuat rating tingkahlaku informan, ia sesuai dijalankan di satu tempat tertutup
- iv. Participant observation : penyelidik turut serta melakukan aktiviti bersama-sama informan.
- v. Direct observation : penyelidik memerhatikan secara langsung informan

Kebolehpercayaan dapatan kajian kualitatif juga akan dicapai apabila kedua-dua *intensive interview* (soalan temu bual disediakan terlebih dahulu) dan *participant observation* atau *semi-structured observation* (*field notes*) dibuat serentak dengan membuat perbandingan dapatan data melalui rating antara dua *independent observers*. Darjah kebolehpercayaan boleh ditingkatkan jika setiap perbandingan telah dibuat pada jarak masa sesuatu sesi sebelum dan selepas⁴⁵. Pada tahap ini, seseorang penyelidik hendaklah melakukan temu bual, diikuti dengan pemerhatian dan diakhiri dengan temu bual semula bagi melengkapkan data-data yang kurang jelas atau kurang lengkap.

Kebolehpercayaan maklumat-maklumat dalam data temu bual dan pemerhatian boleh dicapai semasa penyelidik membuat proses pengkodan. Penyelidik dikehendaki membuat kod ke atas setiap tema atau kategori dan diberikan kepada dua orang penyemak. Kedua-dua penyemak kod tersebut dibekalkan dengan satu salinan garis panduan mengekod berserta jadual kod yang mencukupi. Kedua-dua pengekod tersebut mestilah sepakat dalam memahami kod dan maksud tema atau kategori tersebut⁴⁶. Sebagai contohnya penyelidik telah melantik dua orang penyemak kod yang terdiri daripada:

1. Pn. Norleela binti Syarif (M.A, Pengajian Islam, UKM)
Jabatan Pengajian Agama Islam dan Pendidikan Moral,
Maktab Perguruan Raja Melewar,
Seremban.
2. En. Saidi bin Muhammad (calon Ph.D, Pengajian Islam, UM)
Jabatan Pengajian Agama Islam dan Pendidikan Moral,
Maktab Perguruan Islam,
Bangi.

vi. Indirect observation : penyelidik memerhatikan dari jauh tanpa disedari oleh informan.

⁴⁵ Ibid, *op.cit.* h. 9.

⁴⁶ Chang Lee Hoon, “ Penggunaan Dilema Vignet Dalam Kajian Penaakulan Moral Di Kalangan Pelajar Tingkatan Empat” dalam Marohaini Yusoff (2001), *op.cit.*, h. 179.

Kebolehpercayaan data-data kualitatif boleh dibuat jika deskripsi (transkripsi) yang terhasil disemak oleh satu panel hakim yang terdiri daripada sekurang-kurangnya dua orang penyelidik bebas yang berwibawa. Sekurang-kurangnya 2/3 daripada data perlu disemak dan dipersetujui. Persetujuan yang dicapai antara panel hakim inilah yang akan menentukan kebolehpercayaan dapatan kajian, teknik ini diistilahkan sebagai *discovery procedure*⁴⁷. Oleh itu, dua panel hakim yang berwibawa telah dilantik bagi tujuan ini, mereka terdiri daripada (sebagai contoh):

1. Dr. Johari bin Talib
Jabatan Pendidikan
Maktab Perguruan Raja Melewar,
Seremban.
2. Dr. Jusoh bin Kadir
Jabatan Pengajian Agama Islam dan Pendidikan Moral,
Maktab Perguruan Islam,
Bangi.

Sebagai kesimpulannya, seseorang penyelidik perlulah menitikberatkan kesahan dan kebolehpercayaan instrumen dan data kajian, kerana ia akan menentukan kesahan dan kebolehpercayaan hasil kajian. Dicadangkan teknik triangulasi (soal selidik, temu bual dan pemerhatian juga dokumen) digunakan sebagai instrumen pengumpulan data bagi menyokong dan mengukuhkan antara satu data dengan data yang lain, dengan harapan kelemahan data yang dipungut melalui soal selidik akan dineutralkan oleh data yang dikumpul daripada temu bual dan pemerhatian atau sebaliknya, di

⁴⁷ Wan Zah Wan Ali, "Fenomenografi Dalam Penyelidikan Pembelajaran" dalam Marohaini Yusoff (2001), *op.cit.*, h. 334.

Discovery procedure bermaksud, hasil dapatan yang sama diperoleh oleh beberapa orang pengkaji terhadap sesuatu fenomena. Iaitu menemui pelbagai cara dan mengalami sesuatu pengalaman seumpama menemui spesis tumbuhan baru di sebuah pulau. Penemuan yang sama tidak semestinya dapat diulang di tempat lain. Setelah seseorang penyelidik dapat mengenal pasti ciri-ciri spesis tersebut, maka ia akan dapat membantu penyelidik lain untuk mengesahkan kehadirannya. Ertinya jika seseorang penyelidik memperincikan kategori deskripsi yang terhasil, maka mereka boleh menggunakannya untuk menghakimkan kehadiran kategori dan ciri-ciri dalam setiap subjek kajian yang lain.

sampling penggunaan sampel yang bersifat heterogen yang mewakili seluruh Malaysia.

Rujukan

- American Psychological Association (1985), *Standards for Education and Psychological Testing*. Washington D.C., : APA.
- Bailey K.D., (1978), *Methods of Social Research*. New York : Free Press.
- Borg, W.R., and Gall, M.D., (1979), *Educational Research ; An Introduction*. 3rd Edition, New York : Longman.
- Business and Management*. New York : Mac Millan Publishing, Company.
- Campbell, D., and Fiske, D., (1959), “ Convergent and Discriminant Validation”, *Psychological Bulletin*, Vol. 56.
- Campbell, D., (1955), “ The Informant in Qualitative Research”, *American Journal of Sociology*, vol, LXI , Nos 1-6, July 1955 – May 1956, Chicago, Illinois : University of Chicago Press.
- Chang Lee Hoon, “ Penggunaan Dilema Vignet Dalam Kajian Penaakulan Moral Di kalangan Pelajar Tingkatan Empat” dalam Marohaini Yusoff (2001), *Penyelidikan Kuantitatif, Pengalaman Kerja Lapangan Kajian*. Kuala Lumpur : Universiti Malaya.
- Coleman, J., (1958), “Relational Analysis : The Study of Social Organization with Survey Method”, *Human Organization*, vol. 17.
- Conoley J.C., and Kramer J.J., (1989), “The Tenth Mental Measurements Yearbook” . Lincoln, NE : University of Nebraska Press dalam L.R. Gay dan Diehl P.L (ed) (1992), *Research Methods for Business and Management*. New York : Mac Millan Publishing. Company.
- Conrley, J.C., dan Impara, J.C., (1995), *Thirteenth Mental Measurements Year Book*. Lincoln : University of Nebraska Press.
- Fraenkel, J.R., and Wallen, N.E., (1996), *How to Design and Evaluate Research*. USA : Mc. Fraw-Hill Inc.
- Gay, L.R., (1992), *Educational Research : Competencies for Analysis and Application*. 4th Edition, Maxwell Macmillan International Edition, New York : Maxwell Macmillan Publishing Company.

- Kvale, S., (1997), *The Social Construction of Validity*. Kerja-Kerja European Research of Teaching and Learning, Greece, h. 6-7, dalam Marohaini Yusoff (2001), *Penyelidikan Kualitatif : Pengalaman Kerja Lapangan Kajian*. Kuala Lumpur : Universiti Malaya.
- Kayser, D.J. and Sweetland, R.C., (1990), "Test : A Comprehensive Reference for Assessment in Psychology", Education and Business, 3rd ed., Austin, TX : PRO-ED dalam L.R. Gay dan Diehl P.L (1992), *Research Methods for Business and Management*, New York : Mac Millan Publishing, Company.
- Kayser, D.J. and Sweetland, R.C., (1985), *Test Critiques*, Vol. III. Kansas City : Test Corporation of America.
- Kubiszyn, T., and Borich, G., (1993), *Educational Testing and Measurement*. New York : Harper Collins.
- Kubiszyn, T., and Borich, G., (2000), *Educational Testing and Measurement : Classroom Application and Practice*. Sixth Edition. New York : John Wiley & Sons, Inc.
- Laine, M.D., (2000), *Fieldwork, Participation and Practice : Ethics and Dilemmas in Qualitative Research*. London : Thousand Oaks, California : Sage.
- Lyman, H.B., (1986), *Test Scores and What They Mean*. Fourth Edition, New Jersey : Prentice Hall Englewood Cliffs.
- Mak Lau Fong (1981), *Evaluating Validity and Reliability of Selected Qualitative Research Methods*. Singapore : National University of Singapore.
- Maaneen, J.V., (1983), *Qualitative Methodology*. Beverly Hill, California : Sage Publication.
- Mc Call, G.J., and Simmons, J.L., (1969), *Issues in Participant Observation : A Text and Reader*. Reading Mass : Addison-Wesley Pub. Co.
- Merton, K. and Kandall, P., (1946), "The Focussed Interview", *American Journal of Sociology*, vol. 51., Nos 1-6., July 1945 – May 1946, Chicago, Illinois : University of Chicago Press.
- Miles, M.B., and Huberman, M.A., (1994), *Qualitative Data Analysis : A Sourcebook of New Methods*. Bayerly Hills : Sage Publication.
- Mohd. Salleh Abu dan Zaidatun Tasir (2001), *Pengenalan Kepada Analisis Data Berkomputer SPSS 10.0 For Windows*. Kuala Lumpur : Venton Publishing.
- Mohd. Majid Konting (1993), *Kaedah Penyelidikan Pendidikan*, c.2, Kuala Lumpur : Dewan Bahasa dan Pustaka.

- Osterlind, S.J., (1992), *Constructing Test Item*. 2nd Edition, London : Kluwer Academic Publishers.
- Popham, J., (1990), *Modern Educational Measurement. A Practitioner's Perspective*. 2nd Edition, New Jersey: Prentice Hall, Englewood Cliffs.
- Pratt, D., (1980), *Curriculum Design and Development*. International Edition, USA : Halcourt Brace Jovonarich.
- Robert, J. Mc Dermott and Sarrel, P.L., (1996), *Health Education Evaluation and Measurement : A Practitioner's Perspective*. 2nd Edition, New Jersey : Prentice Hall Englewood.
- Roid, G.H., and Haladyna, T.M., (1982), *A Technology for Test-Item Writing*. New York : Harcourt Brace Jovonovich Publishers.
- Scout, C., (1961), " Research on Mail Surveys", *Journal of the Royal Statistical Society*, vol. 124, Series A (General), London : Royal Statistical Society.
- Sheridan, J. C. dan Steed, L.G., (2001), *SPSS Analysis Without Anguish*. Brisbane : John Willey & Sons Ltd.
- Stake, R.E., (1968), "Toward a Technology for the Evaluation of Educational Program" dalam Ralph W. Tyler, Robert M. Cagne & Michael Scriven (1968), *Perspectives of Curriculum Evaluation*. Chicago : Rand Mc Nally & Company.
- Suseela Malakolunthu, " Pengumpulan dan Analisis Data Kualitatif : Satu Imbasan", dalam Marohaini Yusoff (2001), *Penyelidikan Kualitatif : Pengalaman Kerja Lapangan Kajian*. Kuala Lumpur : Universiti Malaya.
- Vales, R., (1950), *Intercarion Process Analysis*. Reading, Ma : Addison-Wesley.
- Vierra, A. and Pollock, J. (1992), *Reading Education Research*. Scottsdale : Gorsuch Sewishirch.
- Wainer, H., and Braun, H.I., (1988), *Test Validity*. New Jersey : Lawrence Erlbaum Associates Publishers.
- Wan Zah Wan Ali, " Fenomenografi Dalam Penyelidikan Pembelajaran" dalam Marohaini Yusoff (2001) *Penyelidikan Kualitatif : Pengalaman Kerja Lapangan Kajian*. Kuala Lumpur : Universiti Malaya.
- Webb. E.J., Campbell, Schwartz. D.T., R.D and Sechrest., (1966), *Unobtrusive Measures*. Chicago : Rand Mc. Nally.
- Wiersma, W., (2000), *Research in Education : An Introduction*. Boston : Allyn and Bacon.