

Lecture 6

Issues in evaluation
or
lies, damned lies and statistics

Seeking information

- validity & reliability
- kinds of information sought
- sampling
- interviews
- questionnaires

Can we trust it?

- we seek information to
 - inform our design decisions
 - check whether specifications have been met
 - evaluate usability
- no good if it's poor information
- testing/evaluation must be
 - valid
 - reliable

Validity & reliability

- concerned with effectiveness of test/evaluation methods
 - i.e. concerned with *truth*
- evaluation is *valid* if
 - it measures what it claims to measure
- evaluation is *reliable* if it is consistent
 - internally consistent
 - different sections/questions produce consistent results
 - externally consistent
 - repeated evaluations produce consistent results (all other things being equal)

Validity example - the SPOC form

- Student Perception of Course
- supposed to collect valid data on all aspects of students' experience
- tutors get feedback in the form of statistics
 - 90% of students on GIC p/t thought the creche provision was highly satisfactory

BUT GIC p/t was an evening course - no child care provision

Validity & reliability (2)

- for reliable data we need to
 - ask the right questions
 - of the right people
 - correctly interpret results

Validity & reliability (3)

- if a test is invalid or unreliable
 - it is **completely** invalid
 - no such thing as 'partly valid'
 - all data collected should be ignored/discarded
 - no conclusions should be drawn from the data
 - no action should be taken as a result

Causes of invalidity

- invalid/unreliable results can be from
 - poor sampling
 - poor questionnaire design
 - poor administration of test
 - poor analysis of results
 - bias introduced by other factors

Kinds of information sought

- quantitative
 - fact, measurement
 - relatively easy to get
- qualitative
 - opinion, feelings, attitudes
 - difficult to get valid and reliable measures

Sampling

- need for sampling
- getting a valid sample
 - must be *representative*
- statistical calculation
 - sample size, confidence level, confidence coefficient
- specialist, technical knowledge required
- further reading - Kendall & Kendall

Designing a sample

- determine the *population*
- choose *type of sample*
 - *size, representativeness*
 - convenience sample (who's handy/who responds)
 - purposive sample (a chosen group)
 - random sample
 - simple (raffle tickets)
 - complex

Complex random sample

- systematic
 - every *i*th
- stratified
 - sample from each sub-population
- cluster
 - a specific sub-group

Interviewing

- numbers must be limited
 - very time-consuming
 - cover all levels/all functional areas
- clear plan – structured interviews
- recording
 - notes
 - video

..

...

Questionnaires

- validity
 - does it measure what you intended?
- reliability
 - does it give consistent results?
 - questionnaires must be particularly well designed
 - there is no help or feed-back for respondent

..

...

Question design

- types of question
 - open/closed questions
- introducing bias
- methods for getting & quantifying answers
- attitude scales

..

...

Open questions

- easy to design
- respondent provides rich and varied answers
- difficult to evaluate

..

...

Examples - open

What are the most frequent problems you experience with computer output?

Which of these is the single most troublesome?

Why do you think this is?

..

...

Closed questions

- more difficult to design
- limited range of responses
- more easily quantifiable

..

...

Examples - closed

The sales figures are usually late.
 Agree Disagree

When the sales figures are prepared by computer data services they are late.
 Never Rarely Sometimes Often Always
 1 2 3 4 5

..

...

Question design - neutrality & leading questions

Isn't there a better way to project sales than the antiquated one you're using now?

You agree with other managers that inventory control should be computerised, don't you?

What do you think of computerised inventory control?

..

...

Using response scales

- used to measure attitudes
- get respondents' judgments about a topic
- four types of measurement scale
 - nominal
 - ordinal
 - interval
 - ratio

..

...

Nominal scale

What type of program do you use most?

1. Word Processor
2. Spreadsheet
3. Database
4. Graphing program

Can get totals for each classification. No ranking is assumed.

..

...

Ordinal scale

- allow classification with rank order

The support staff is:

1. extremely helpful
2. very helpful
3. moderately helpful
4. not very helpful
5. not at all helpful

..

...

Interval scale

How useful is the support given by staff in the information centre?

1 2 3 4 5
 Not at extremely
 all useful useful

Here we assume the intervals are equal

..

...

Ratio scale

Approximately how many hours do you spend on the computer daily?

0 2 4 6 8

Used when intervals are equal, and there is an absolute zero starting point

..

...

Some references

- Kendall & Kendall
- Faulkner
- B White Dissertation Skills, pub Cassell 2000 Ch 4 & 5
- www.mors.org/education_colloquium/EC2000/presentations/polakoff

..

...

Evaluation of questionnaires

- should be trialled and validated before use
- check for appropriate *language*
- appropriateness of *scales*
- ability to extract and analyse useful data

..

...

Summary

- evaluation methods must be valid and reliable
 - invalid or unreliable data is worse than no data
- interviews must be carefully planned
- questionnaires must be carefully designed
 - avoid bias
 - use appropriate scales and quantification mechanisms
- results must be quantifiable

..

...