

Quantitative and qualitative data analysis: Issues and opportunities

By Peter Grimbeek

Some quantitative references

- Blaikie, N. (2000). *Designing social research*. USA: Polity Press.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. USA: Lawrence Erlbaum Associates.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS*. London: Lawrence Erlbaum Associates, Publishers.
- Campbell, Donald T., & Stanley, Julian C. (1972). *Experimental and Quasi-experimental designs for research*. USA: Rand McNally & Company.
- Lee, C. (1998). *Alternatives to cognition: A new look at explaining human social behavior*. London: Lawrence Erlbaum Associates, Publishers.
- Meehl, P. E. (1970). Theory testing in psychology and physics: A methodological paradox. In D. E. Morrison & R. E. Henkel (Eds.), *The significance test controversy* (pp. 252-266). USA: Aldine Publishing.
- Michell, J. (1999). *Measurement in Psychology: A critical history of a methodological concept*. United Kingdom: Cambridge University Press.

Some qualitative references

- Freebody, Peter (2003) Qualitative research in education: Interaction and practice, London: Sage.
- Halliday, M.A.K. (1994). Introduction to Functional Grammar, Second Edition, London: Edward Arnold.
- Silverman, D. (2001) Interpreting qualitative data: Methods for analysing talk, text and interaction, London: Sage.
- Silverman, D. (2002) Qualitative research: Theory, method and practice, London: Sage.
- ten Have, Paul (199), Doing conversation analysis: A practical guide. London: Sage.

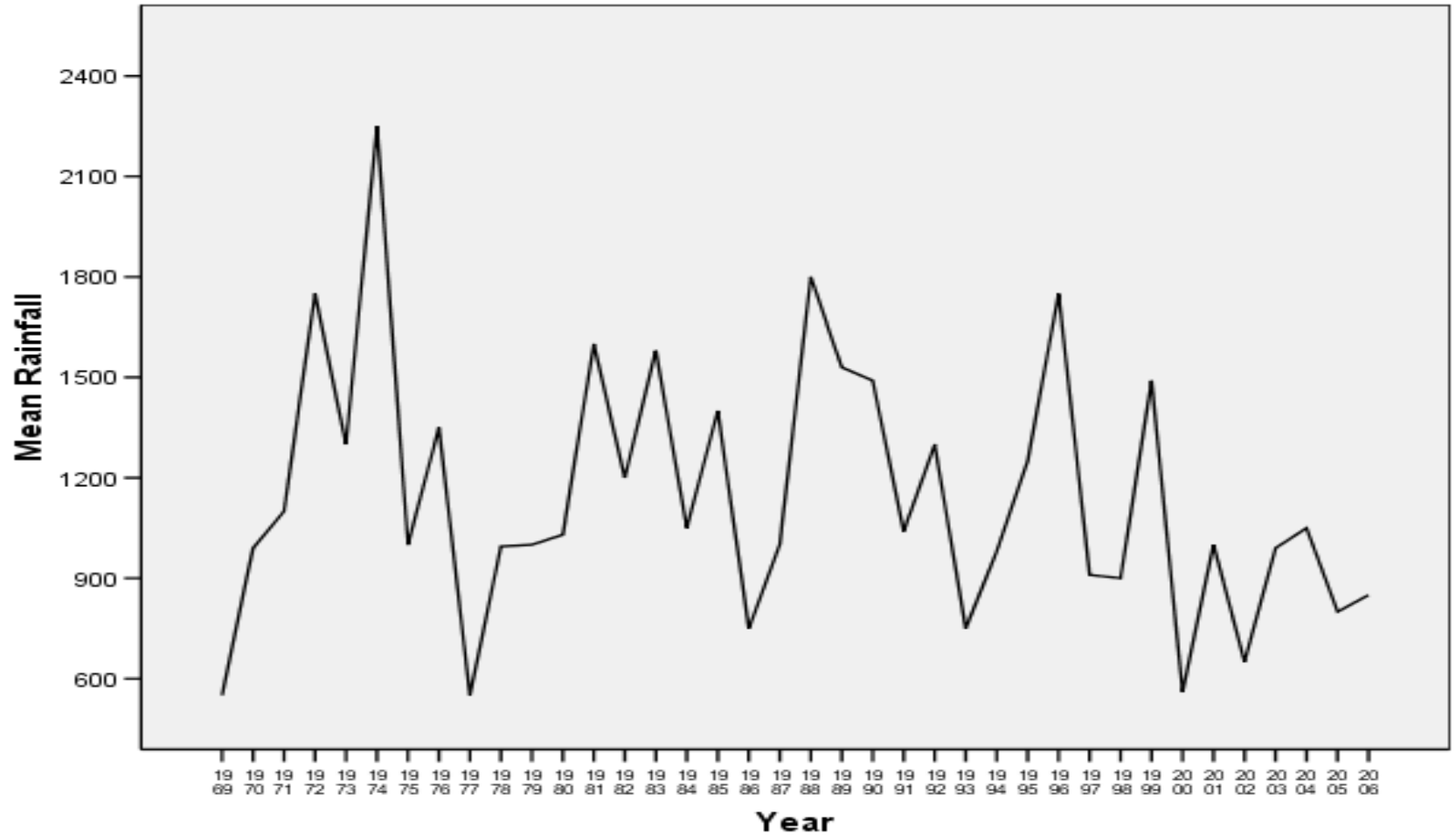
Quantitative and qualitative data analysis: Issues and opportunities

- Work with quantitative data, needs to be cognisant of the following issues:
 - Measurement properties of variables
 - Hypothesis testing
 - Cause-and-effect assumptions

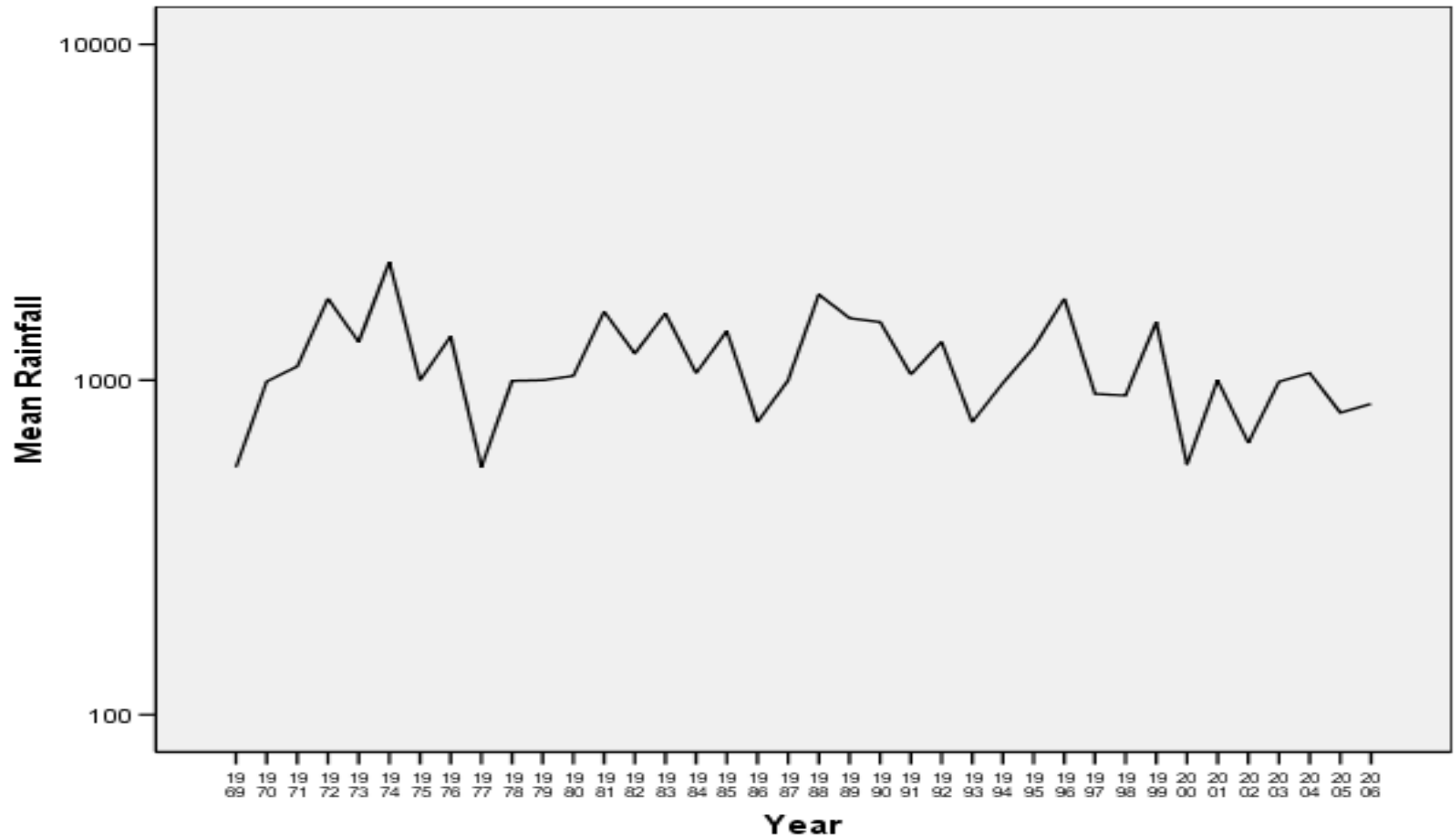
Measurement properties of variables

- The paradigm for quantitative analysis (Michel, 1999) is that variables have parametric measurement properties. That is, data points are equal-interval and normally distributed.
- In reality, most of the data collected for scientific purposes is nonparametric. That is, not only are the distributions skewed and kurtotic but the intervals are non-linear (i.e., exponential or rank-ordered at best).
 - For example, the mean annual rainfall over a period of years is nonlinear in at least two ways (cyclical, exponential) in terms of its rate of change.

Line graph of rainfall over 27 year period



Line graph of rainfall (Natural log transform)



Measurement properties of variables

- Likert response scales provide a graphic example of the difficulties in assigning equal-interval measurement properties
 - Respondents choose between responses conveying, say, Levels of agreement.
 - Qualitative (non-numeric) responses are coded numerically by assigning a number series (usually 1-5 or 1-6).
 - The resulting data set is analysed as if it possesses interval properties.
 - Typically, researchers report the mean score per item, add item scores, etc.

Measurement properties of variables

- Given the choice of assigning Level of Agreement for the following items (from Bond & Fox, 2001):
 - I vomit regularly to control my weight
 - I count the fat grams of the food I eat
 - I exercise solely to burn calories.
- It would be nonsensical to treat these three items as equal contributors to an eating disorder scale.
 - Strong agreement with item 1 indicates a far greater problem than strong agreement with the third item. Also, item 2 is a 'low concern' item with respect to eating issues.

Hypothesis testing

- The business of testing the strength of the association between cause and effect is biased in favour of the alternative hypothesis.
 - That is, almost any hypothesis, and thus almost any cause-effect connection, is susceptible to being confirmed statistically if conditions are right (sample large enough, measure sensitive enough, correct background conditions, etc).
- This phenomenon stems from the commonplace observation that almost any variable produced by human beings is associated at least weakly with every other variable.
 - The apocryphal story here is that with a sample of 50000 or so, all associations turn out to be statistically significant.

Cause and effect

- The notion that we can under proper conditions use quantitative methods to make judgments about cause-effect chains, and having tested an instance of this in a sample, generalise statistically to its relevant population.
- Just as we accept the paradigm of equal-interval variables, so we make the paradigmatic assumption that cause-effect can be determined under proper conditions.

Proper conditions for examining cause and effect

- Ideal conditions are generally assumed to be those of experimental research.
 - Where outcomes are measured under conditions in which at the very least, participants have been assigned to experimental conditions randomly, other conditions are controlled or allowed to vary, with the number of participants sufficient to ensure normalcy of distribution for uncontrolled (individual) differences, and where at best, the sample was drawn randomly from its population.
- In fact, laboratory style experimental work tends to adhere to random allocation, and survey research at its best adheres to random sampling, so that usually one of these conditions is present.

Cause-effect reasoning

- The deductive notion of being able to use quantitative methods to test cause and effect exhibits specific flaws.
- Akin to the process whereby variables are automatically identified as equal interval regardless of psychometric merit, the process whereby variables are identified as causal does not guarantee this to be the case.
- The act of identifying any cognitively based variable (i.e., an attitude, perception, etc) as part of the cause rather than the effect is somewhat arbitrary, and at best as good as the theoretical framework that framed it.

Cause-effect reasoning

- Human data involves testing the relationship between subjective attributes such as self-efficacy and objective attributes such as time to complete a task or success versus failure in completing that task (i.e., speed versus accuracy trade-off).
- Tend to be a range of overlapping conceptualisations of these subjective attributes, operationalised in the form of often highly correlated variables.
- The question is whether the strength and direction of correlation is a guide to cause-effect **explanation** or a numerical indicator of the relationship between overlapping **descriptions** of common attributes.

Cause-effect reasoning

- Christina Lee questions the possibility of postulating cause-effect relationships between subjective attributes and objectives attributes or between subjective attributes and other subjective attributes.
- Lee (& Skinner's) behaviourist perspective:
 - Past behaviour is better guide to present behaviour than past cognitions.
 - Problem in that mental processes are invisible.
 - Competing models may deal with same attributes in different ways.

The end-point of all this doubting

- In sum, deductive strategies when applied to human data encounter obstacles in specifying cause and effect relationships between variables, most particularly when examining associations between objective and subjective attributes or between two or more subjective attributes.

The pragmatic approach

- One option is to be utterly pragmatic and to assume that regardless of origin, variables can be classified as suitable for parametric analysis as long as they are measured along four or more units.
- This approach is analogous to the kind of judgment one might make when estimating the average score.
 - Given that one computes an average score by adding individual scores and then dividing the total by the number of scores, it follows that estimates of the average score become increasingly reliable as the number of individual scores increase.

Skimpy response categories

- It follows that estimates based on less than four scores tend to be unreliable (and also tend to be skewed and kurtotic).
 - So, the dichotomous response scale favoured by instruments such as the Perception of Ability Scale for Students (PASS) (e.g., I am good at printing {No, Yes}) is easy to score manually, and suits in-the-field applications, but is more likely to result in skewed and kurtotic responses.

Generous response categories

- In contrast, when responses are collected across a range of eight to ten points, then not only does a smoother distribution of response become possible but also an average score based on such scores tends to be more reliable.
- A limit to this process is that human beings don't process a large range well.
 - Due to the limits of short-term memory, which posits the rule of seven items, plus or minus two, data collection on response scales that provide between five and ten steps is probably optimal.

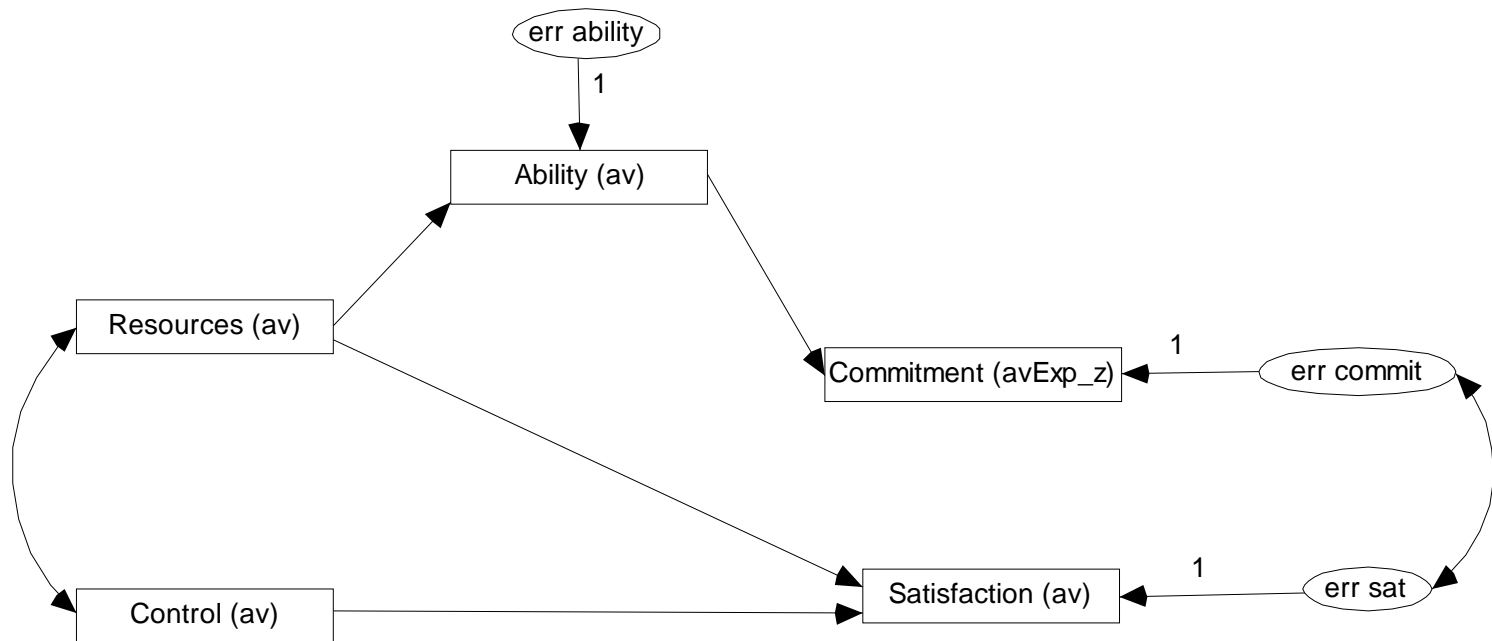
The analogy extends to items, and samples

- More generally, just as responses to items benefit from a perceptually acceptable range (5-10), in the same way, one would expect instruments with more items per scale to produce more reliable results.
- It is probably no accident that scales with 8-10 items tend to do better in terms of reliability statistics (e.g., Cronbach's alpha).
- While the response scale per item and the number of items per scale influence the reliability of estimates, the size of the sample is an even more important consideration.
- In short, size matters.

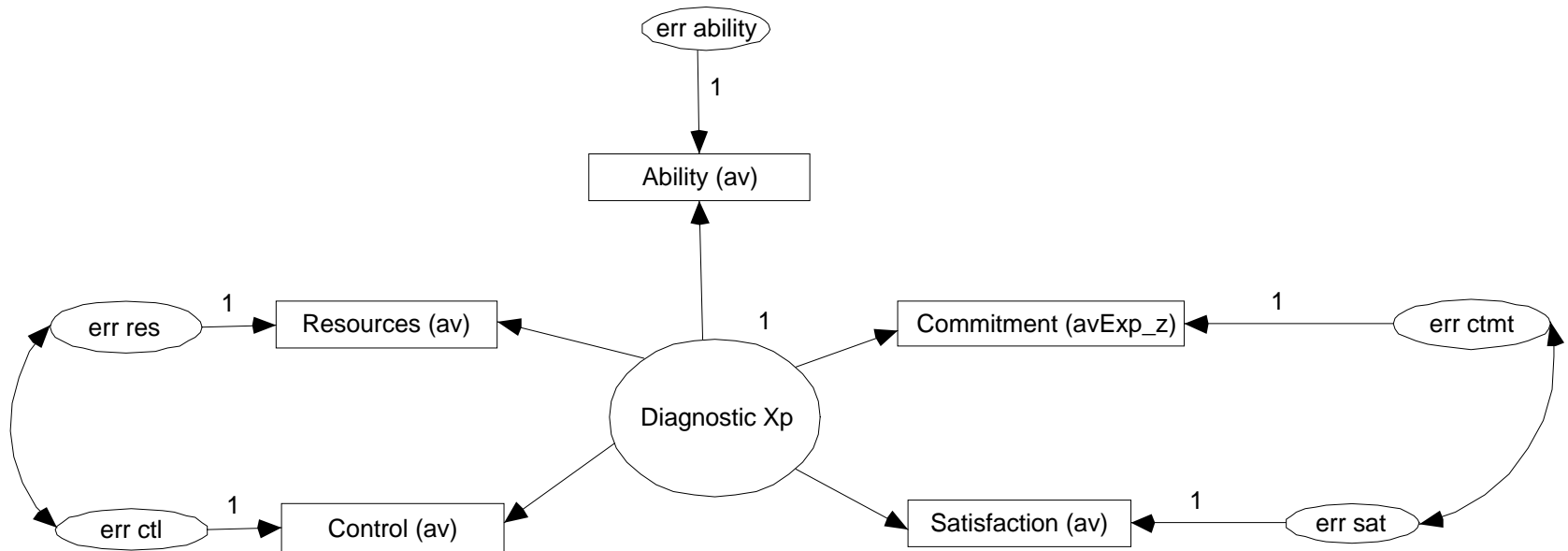
AMOS and reliability

- AMOS uses Maximum Likelihood estimation by default, which assumes parametric psychometric properties, including normalcy of distribution as a sine qua non.
- That aside, one of the charms of AMOS is that it allows one to construct models to estimate a range of analysis simultaneously, which of course enhances the reliability of that analysis.
- AMOS considers both bidirectional and directional effects.
- Its procedures allow factor analysis, regression, and repeated measures to be test simultaneously, as long as the variables meet either the real or pragmatic requirements for parametric analysis.
- In that the model is specified a priori, the notion of cause-and-effect can be tested explicitly via structural equation modelling (SEM).

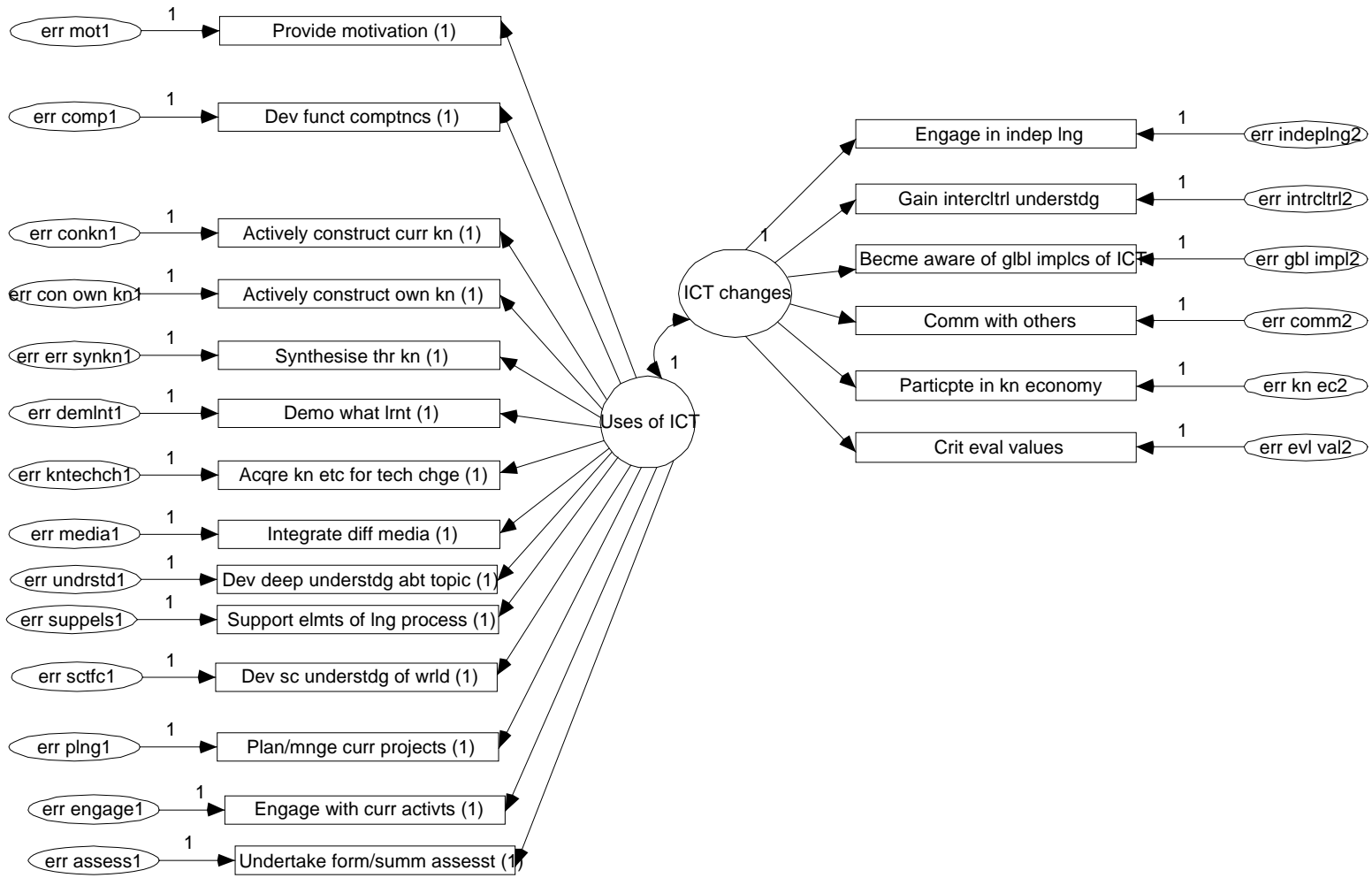
AMOS models



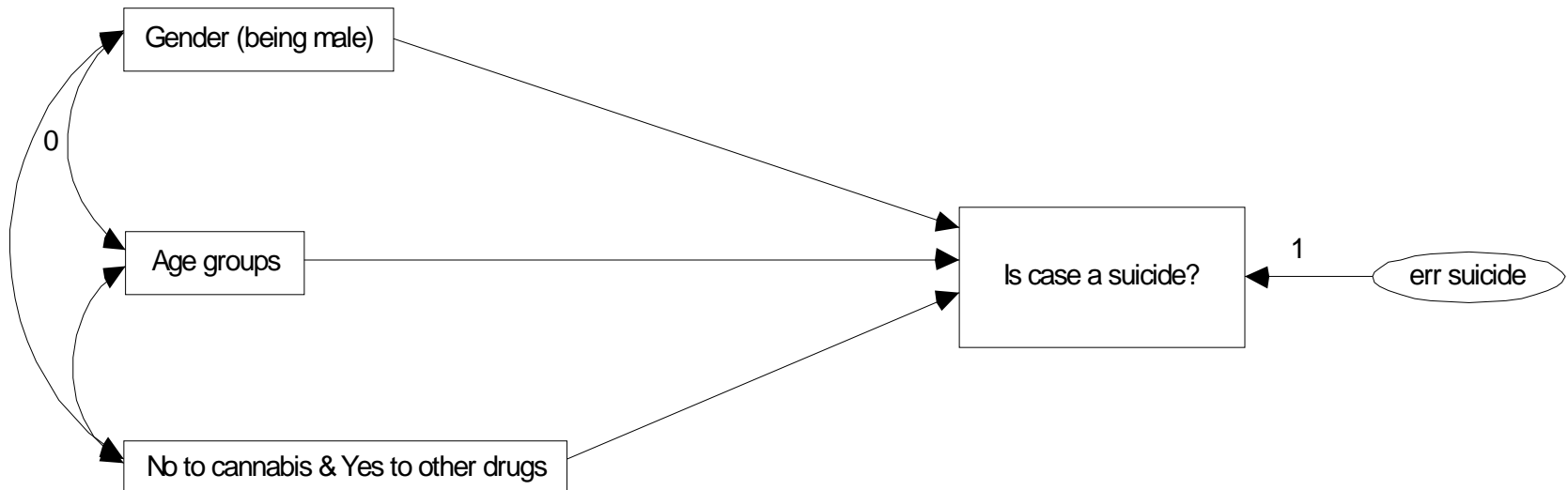
AMOS models



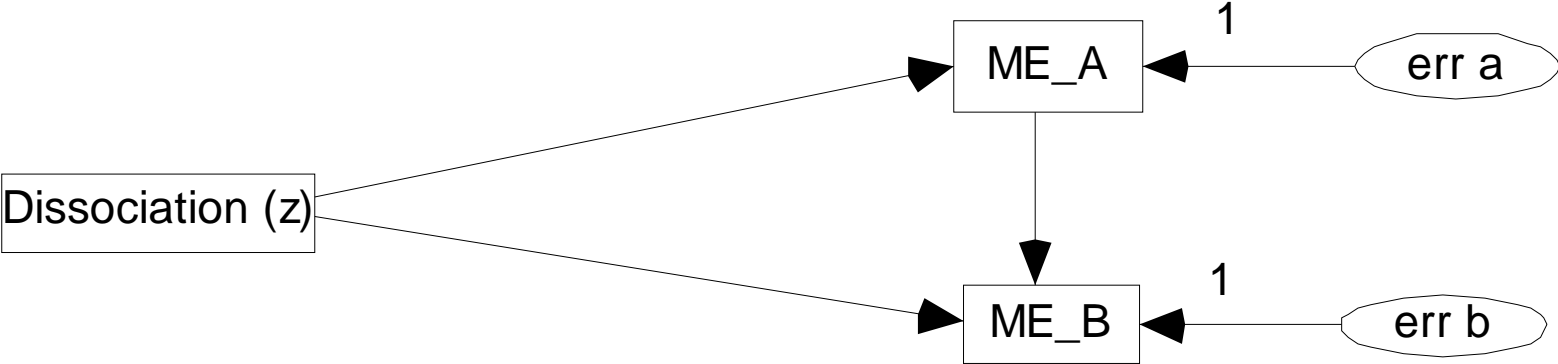
AMOS models



AMOS models



AMOS models



Alternatives

- Putting the joys of AMOS aside, nonparametric tests don't assume the equality of the interval and don't necessarily assume distributional normality.
 - Procedures of interest include the nonparametric equivalent of one-way ANOVA (Kruskal-Wallis), the nonparametric correlation coefficient, Spearman's Rho, and various versions of multiple correspondence analysis (e.g., SPSS Optimal Scaling).
- Optimal Scaling is of particular interest because it allows one to model, say, demographic data, a use for which Pierre Bourdieu, the noted French sociologist created its lineal ancestor, Correspondence Analysis.

Rasch item analysis

- Another alternative is to use Rasch item analysis procedures (e.g., Bond & Fox, 2001, 2007; See your local expert, Rosemary Callingham). Rasch procedures explicitly assume that variables are categorical or ordinal at best, a highly sensible assumption if one is dealing with, say, Likert scale data.
- Rasch software transforms raw data into logits (a form of z-score or log-score with a range of about 3 steps above and below a central point of zero). Rasch software can be used to analyse logit scores both in terms of item difficulty and participant ability.

Using Rasch in everyday life

- The Rasch software, Winsteps, has the charm of being able to read SPSS datasheets and generate Excel output files on request.
- I've found output related to the relative difficulty of Likert scale responses (say, agree vs. strongly agree) quite useful.
- Items can be judged to be disordered when participants in effect rate agreement as more difficult than strong agreement, as opposed to the other way round.
- Rasch analysis clusters items in terms of level of difficulty rather than by default separating them into separate clusters as is the case when using classic methods (e.g., exploratory factor analysis).

Illustration: Three-factor solution for seven items

Factors/Items	1	2	3
School analyses numeracy tests to identify topics causing difficulties (Q8)	0.92		
School analyses numeracy tests to identify gaps in content taught (Q7)	0.89		
School analyses numeracy tests to identify pupils with difficulties (Q6)	0.78		
Use numeracy tests to give students feedback on strengths (Q19)		0.91	
Use numeracy tests to inform students about weaknesses (Q20)		0.90	
Numeracy tests indicate quality of school's numeracy program (Q3)			0.80
Numeracy tests indicate teacher's ability to teach mathematics (Q2)			0.80

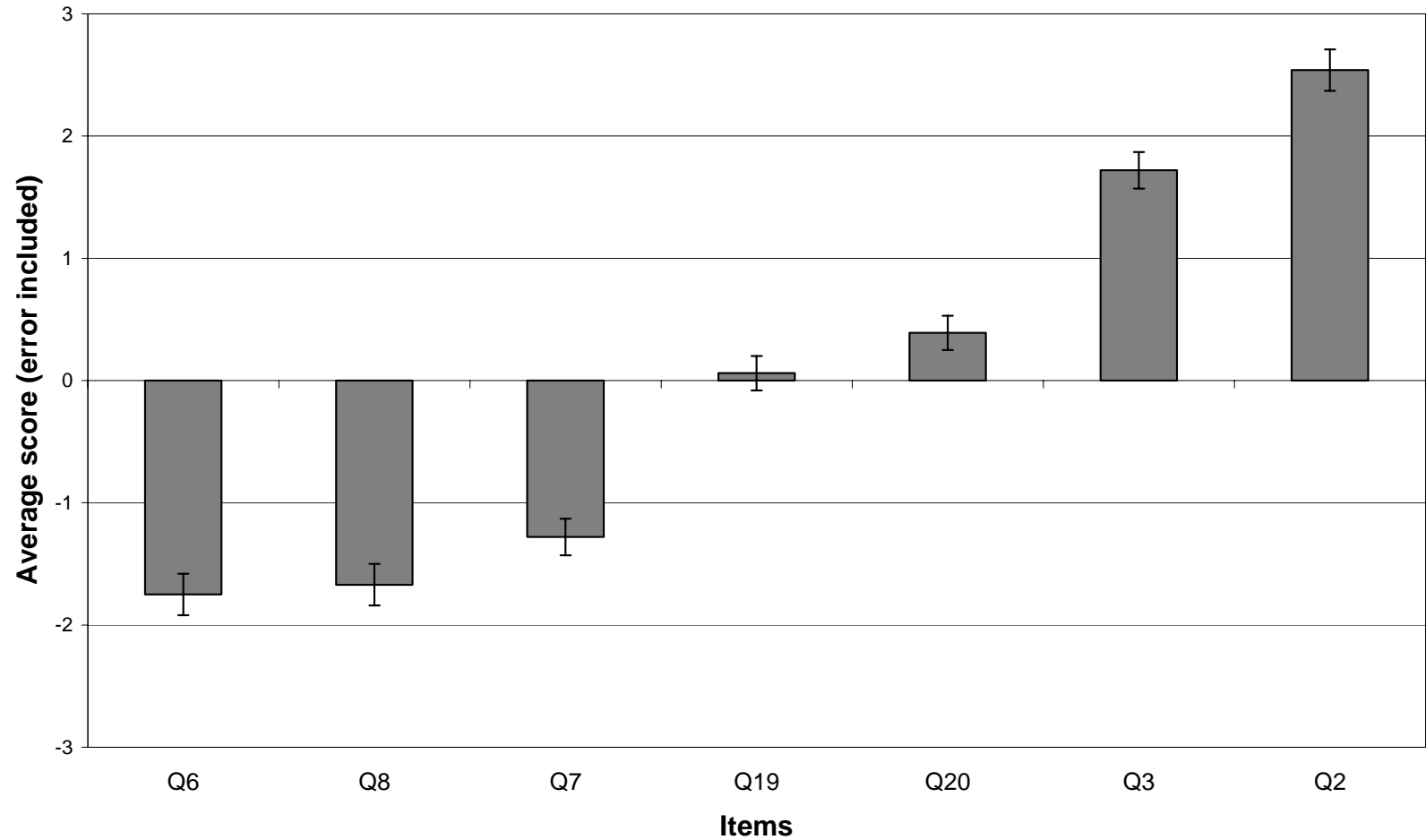
Example of disordered response categories identified by using WINSTEPS

Strongly Disagree

Item	Code (Response category)	Average score
Q4	1	-1.85
Q4	2	-0.22
Q4	3	-0.64
Q4	4	-0.24
Q4	5	-0.06

Strongly Agree

Average score per measure in logits for 7 items



Summary to this point

- To summarise, quantitative analysis has issues related to the measurement properties of variables, hypothesis testing, and cause-effect reasoning, which can be dealt with by resorting to nonparametric analytic procedures or Rasch item analysis or by proceeding cautiously in the full knowledge of the limitations of one's practice.

Qualitative analysis

- The analytic processes associated with qualitative analysis, whether in relation to the analysis of transcripts of conversations or interviews or sections from documents, are less standardised than those for quantitative analysis.
- Rationale for this includes the relative diversity of research strategies.
- Can distinguish three major qualitative research strategies (inductive, retroductive [critical], abductive [interpretivist]), roughly aligned with broad distinctions between:
 - Positivist vs. nonpositivist research strategies (methodologies)
 - Objective (outsider) vs. subjective (insider) accounts

Positivist strategy: Induction

- Inductive researcher assumes that reality can be observed directly via the senses and that the senses suffice to identify uniformities.
- By definition, inductive reasoning proceeds from particular to general
- In theory, inductive research commences with observation but, in practice, observation is prefaced by deciding what to observe (i.e., variables operationalised based on theory).

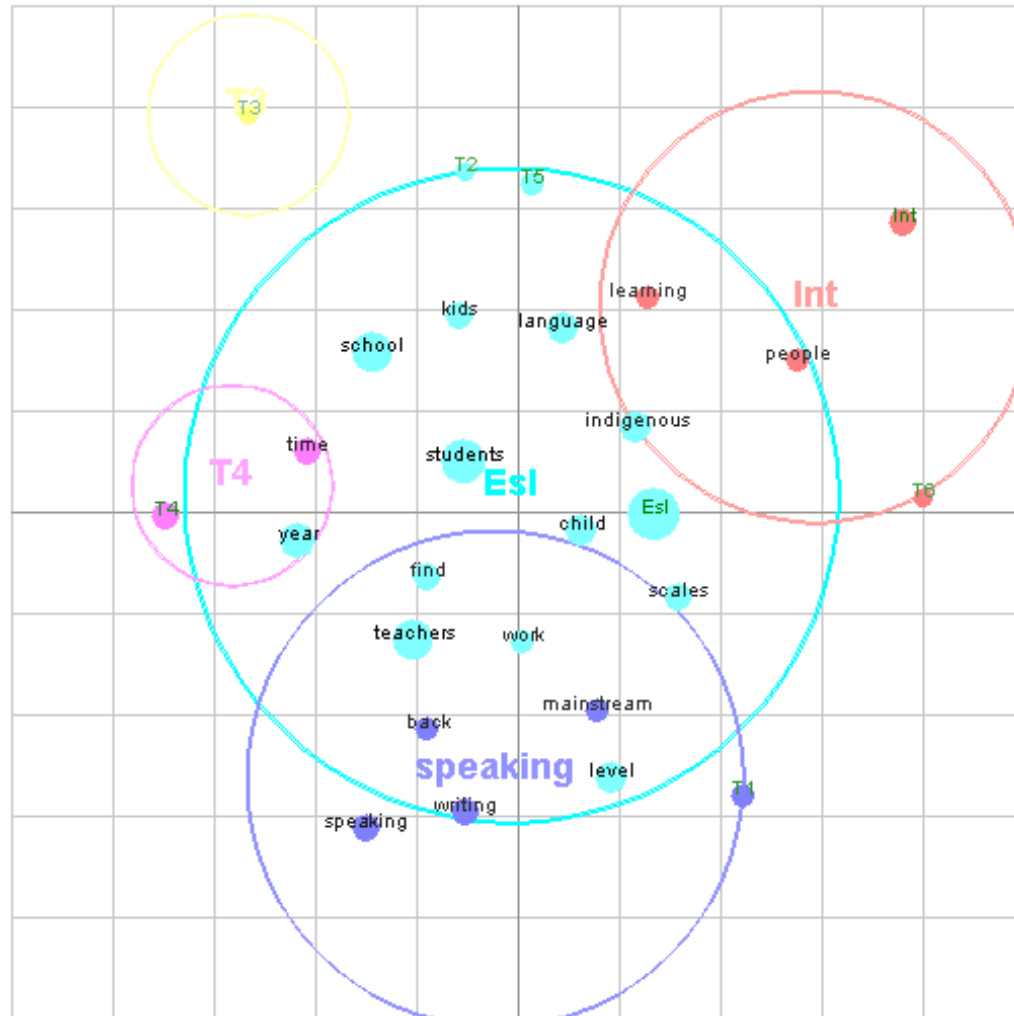
Nonpositivist strategies

- Nonpositivists abandon the view that reality is uniform and orderly (ontology) or that one can independently gain generalisable knowledge about it (epistemology).
- Various (postmodern) schools of thought based on one or both of these assumptions can be identified, including: Constructivist, critical, cultural, scientific realism, retroductive, interpretivist, and abductive.
 - One group of nonpositivist strategies assumes reality is uniform and orderly but not entirely directly observable (by researcher or by others). Implicit assumption that generalisations generated by objective accounts overlook inherent tensions in multiple viewpoints of social actors (e.g., class struggle)
 - Interpretivist research strategy makes more radical assumption that notion of uniform reality is itself a social construction, and while social actors might construct realities cooperatively, they also construct distinct realities.

Theoretical framework versus research strategy

- An issue here is that qualitative researchers tend to combine theoretical frameworks that emphasise a nonpositivist approach with positivist content analysis.
- So, an ESL researcher, with an interest in the extent to which classroom teachers use the ESL bandscales as a guide to classroom teaching, might posit tensions in classroom settings but in fact limit his/her examination of the transcript to responses related to that all important topic, the ESL bandscales as an educational tool.

Illustration: ESL research



Quant and qual analysis issues

The importance of content analysis

- Regardless of the theoretical framework or research strategy, for that matter, meaningful qualitative analysis begins by specifying categories or by counting frequencies of certain kinds of words
 - Count instances of specified categories
 - Specify categories based on frequency of terms
- Issues related to content analysis:
 - Identifying a sufficiently precise set of categories
 - Specifying categories reliably (similar outcomes by different researchers and validity (sensible outcomes))
- Content analysis can be end-point of an analysis of data such as responses to open-ended survey questions

Manual methods

- A related issue is that content analysis is either undertaken by using largely manual methods.
- These include:
 - Coding in Word document
 - Making notes based on documents spread across garage floor
 - Database style instruments such as NUD*ist or NVivo that allow one to tag bits of the transcript or other document in terms of preconstructed tree nodes or as additional free nodes, etc.
 - Despite the increasingly slick interface, and a Word-like capacity to search text for words or groups of words to be tagged as tree nodes or added as free nodes, this kind of software offers a fairly dumb interface.

Sampling issues

- In the end, regardless of whether one codes on the garage floor or within the electronic window provided by NVivo, it is difficult to code more than a smallish sample efficiently.
- This deficiency in qualitative analysis gets dressed up in terms of the art of selecting theoretically relevant passages, etc, but remains a deficiency nonetheless.

Enter left, Leximancer

- Leximancer is automated software that addresses the issue of the starting point and the issue of sample size, and in doing so shifts qualitative analysis from being a form of handcraft for the gifted to being a form of scientific research.
- The revolution in qualitative methods that it provides is equivalent to the shift from laying out books by placing bits of lead in a printing press to laying out books by using software and a laser printer.
- In short, a far more sophisticated style of writing is possible when one uses a word processor.

Leximancer generates a thesaurus

- Leximancer first generates a thesaurus of terms based on the documents to be analysed.
- Then it uses a preset list to exclude semantically uninteresting terms such as, “a” and “the”, brackets, and numbers.
- It scans through the remaining terms, looking for centrally located terms or archetypal concepts.

Leximancer identifies frequently used concepts

- The software uses a clumping algorithm from physics to identify clumped terms and to identify the terms that are most centrally located within these clumps, and compiles a list of these frequently used 'concepts'.
 - For instance, the concept of dog is generally accompanied by words such as tail, fleas, and fur.

Leximancer identifies co-occurrences

- Having identified such a set, the software now examines the extent to which these concepts do or don't co-occur.
- It does so by coding their presence or absence in a series of three-line (a default setting that can be varied) windows.
- The resulting (0,1) matrix represents the extent to which any two variables co-occur, an outcome that is represented visually within the two-dimensional space of a concept map.

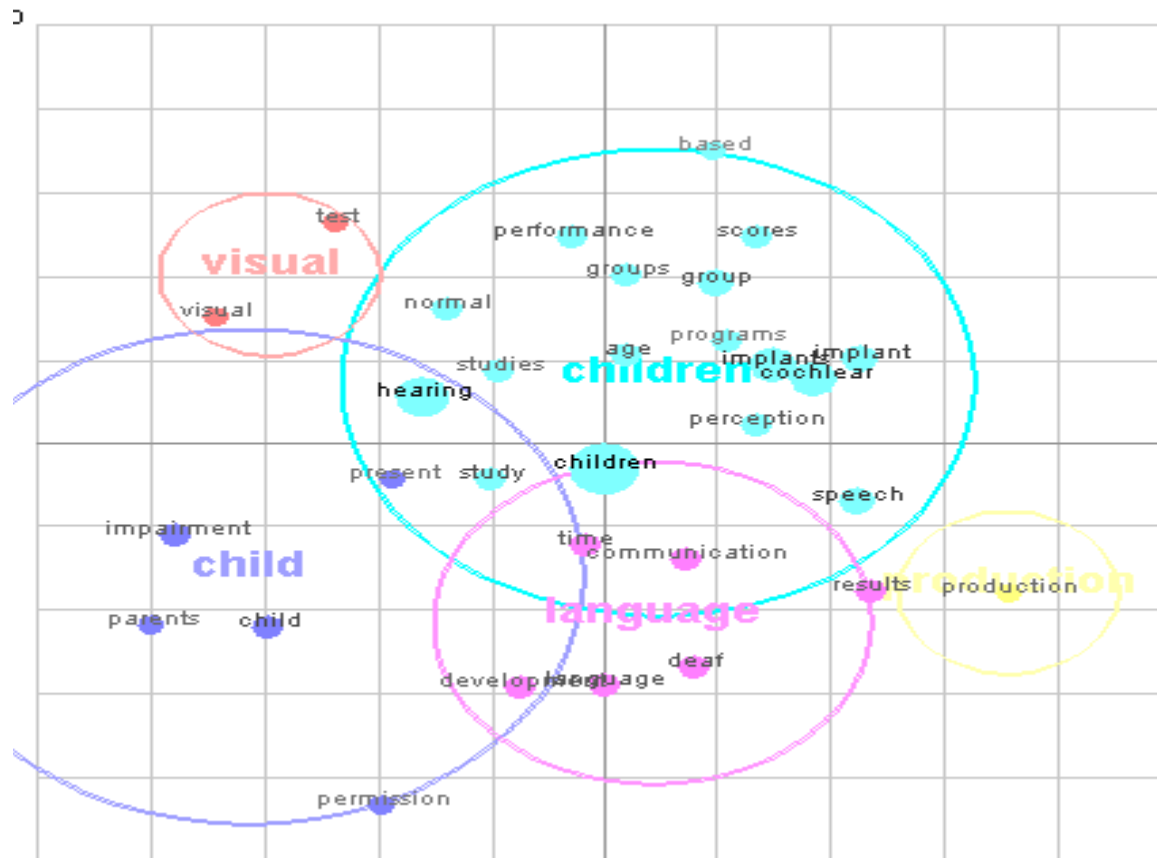
Leximancer produces a concept map

- The concept map depicts the concepts generated from that initial list of terms within a four quadrant space that quantifies their spatial co-occurrence or separation (a bit like cluster analysis).
- This concept map provides the starting point for a higher-order analysis based on the relative positions of clusters of concepts.

Leximancer data sets

- Leximancer can use data stored in Word, PDF, and a range of other file formats.
- It can be used to analyse a wide range of data, including:
 - Conversation transcripts (ref., ESL focus group)
 - Entire literatures (ref., deafness related to hearing aids)
 - Dissertations (ref., Internationalisation of universities)
 - Test instruments (ref., Two emotional intelligence tests)

Analysis of deafness literature related to cochlear implants



Analysis of dissertation on topic of internationalisation

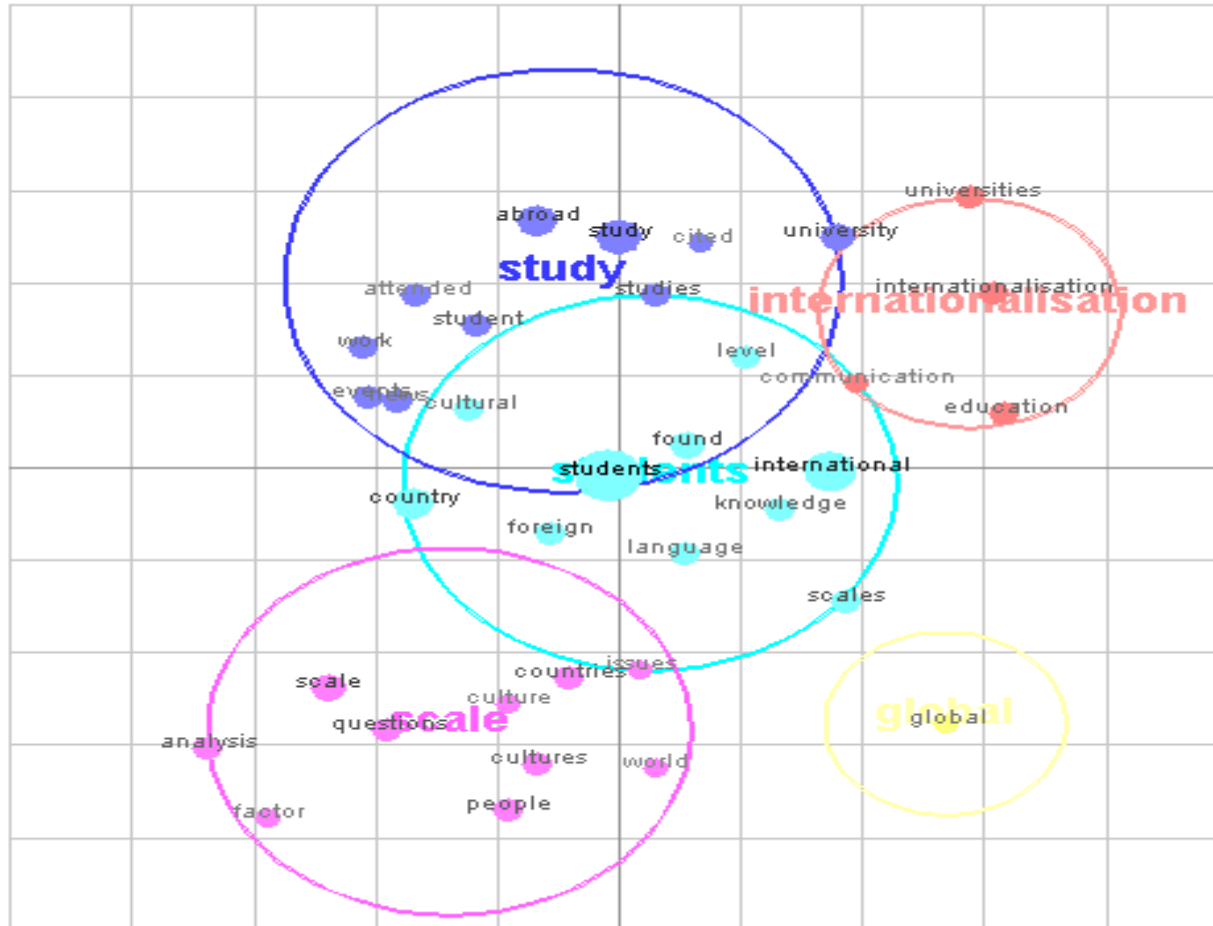


Illustration: Emotional intelligence items (Thai EQ test)

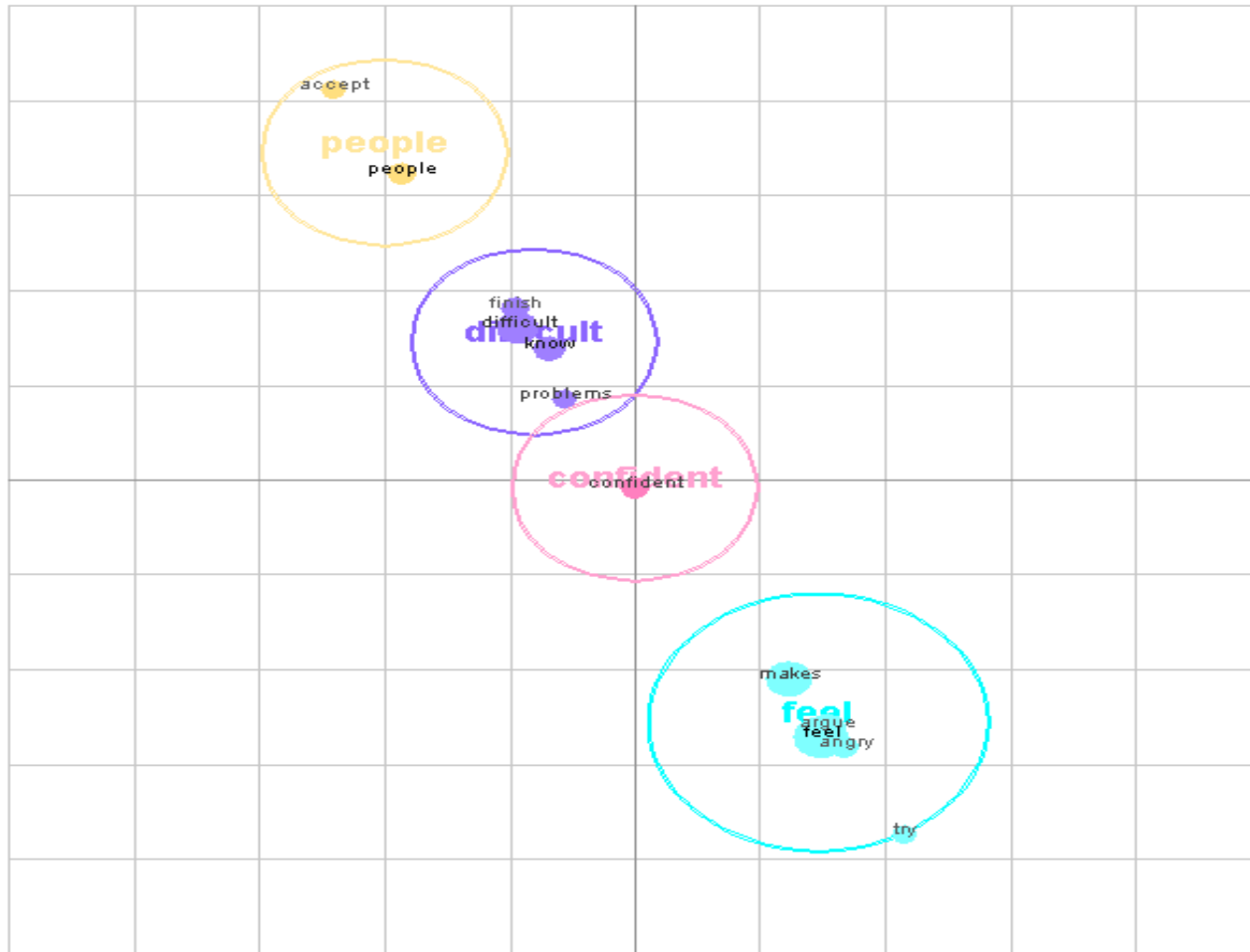


Illustration: Emotional intelligence items (Wong & Lau test)

