



Introduction to Scale Development

Dr Mark Griffin
ADASIS Consulting
www.adasis-oz.com

Motivating example

- The Queensland Department of Communities, Child Safety and Disability Services is currently rolling out the Positive Parenting Program across Queensland
- Triple P aims to increase parenting skills and subsequently child wellbeing
- My company is evaluating the success of this program



Motivating example

- Within such an evaluation how can we possibly measure a construct as complex as parent or child wellbeing?
- One answer is through the use of scales such as the DASS21 (Depression Anxiety Stress Scales)
- Within the DASS21 each participant is asked to complete 21 questions on a sliding scale
- I found it hard to wind down – never, sometimes, often, almost always



Motivating example

- The responses to these 21 questions are then coded in such a way that we obtain a single number for each person according to a number of domains
 - Depression
 - Anxiety
 - Stress
- And an overall number representing wellbeing

In this session

- Designing good questions for a scale
- What is the difference between exploratory and confirmatory factor analysis?
- What is Structural Equation Modelling?
- What is Rasch analysis?
- What does reliability and validity mean?
- Adapting an existing scale to a new construct or a new population



Target audience for this session

- This session is designed for a target audience of evaluators who are new to the field of scale development.



Designing questions for a scale

- There are two aspects that we need to consider when designing questions for a scale
- We want questions that are easily understood and clear to answer (whether it be a survey or otherwise)
- We want to make sure that each question is related to the underlying dimension we want to measure (eg. wellbeing) ... we will come back to this at the end of this talk when we discuss reliability



Designing questions for a scale

- Tourangeau (2000) identifies seven types of misinterpretation:
 - Grammatical ambiguity (eg. How often do “you” hug your child?)
 - Excessive complexity
 - Faulty presupposition
 - The question makes an assumption about the state of the world, and then asks the participant about this assumption (eg. How do you feel when you hit your child?)



Designing questions for a scale

- Vague concepts
 - What is a child?
- Vague quantifiers
 - How often is “often” ?
- Unfamiliar terms
 - Why use the word “clinch” when you could just say “hug”



Designing questions for a scale

- False inferences
 - Participants make an assumption about what the question is trying to ask.
 - They then have two possible questions – the literal interpretation, and the assumed interpretation.



Scoring your participant's response

- There is a lot that we can say about questionnaire design, but let's suppose that we now have a complete set of questions/responses.
- The next thing that we will do is to combine these responses into a single metric
- The most common technique at this stage is known as exploratory factor analysis (or in most cases simply factor analysis)

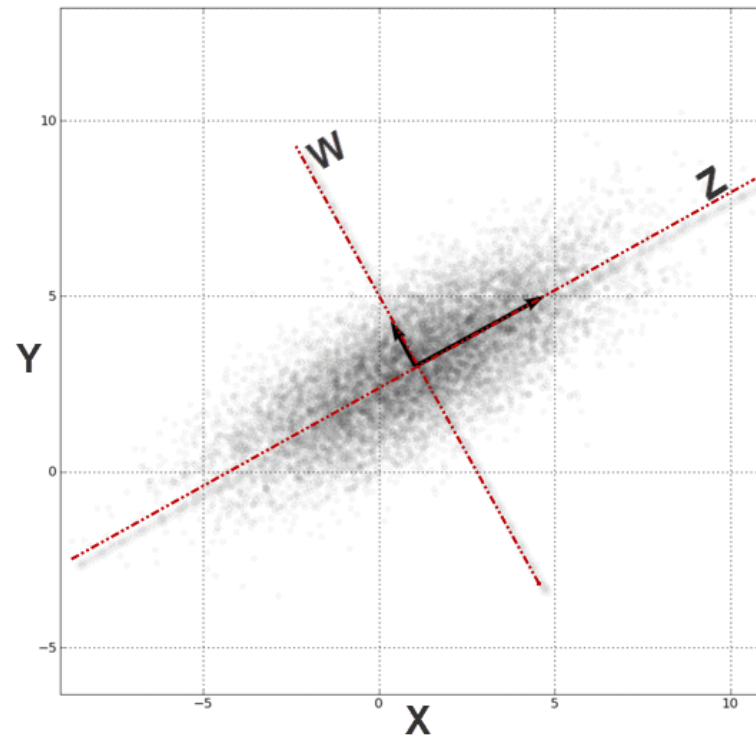


Factor analysis

- For simplicity let us start by considering a scale with just two questions
- There will be some similarities in how participants answer each of these questions (questions which are about depression might obtain similar responses)
- There will also be some differences in these responses (depression might mean different things to different people)

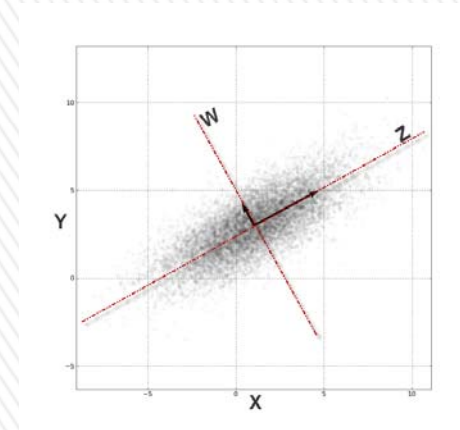


Factor analysis



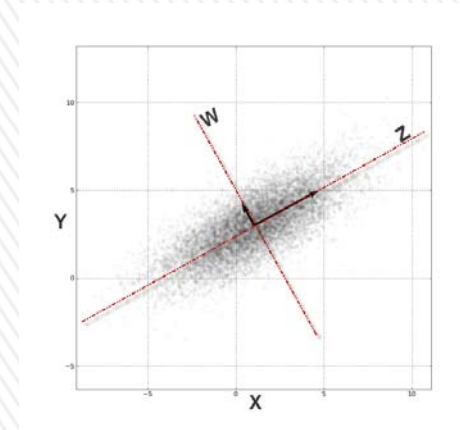
Factor analysis results

- Factor analysis produces a set of results
 - Factors
 - Eigenvalues
 - Factor loadings



Factor analysis results

- Factors describe the relationship between the new dimensions and the original dimensions

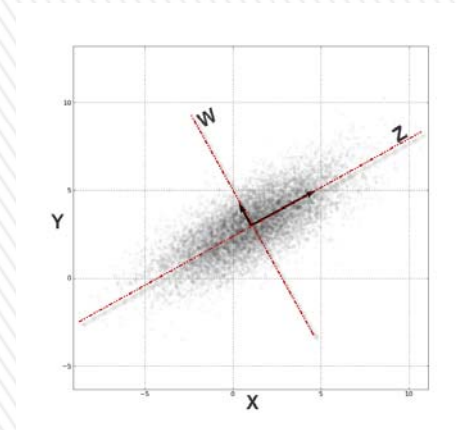


- In our case we have two factors – Z and W
- Eg. $Z = 0.3 X + 0.4 Y$



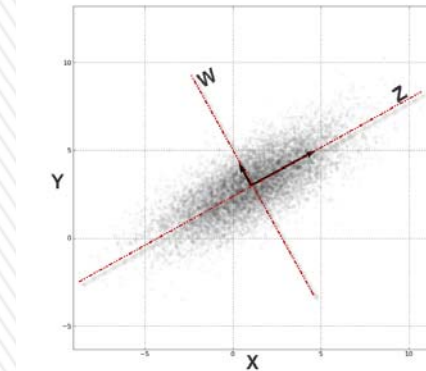
Factor analysis results

- Eigenvalues – we have an eigenvalue for each factor which indicates how much variability in the data can be explained by that factor
- Factor loadings – for each person who completed the survey we now have a score for that person against each factor



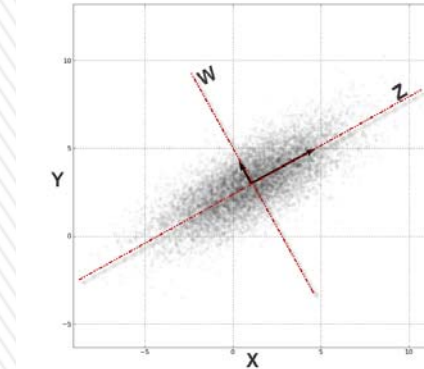
Measured and latent variables

- Within such a factor analysis we can say that we have a set of measured and latent (or hidden) variables
- We have measured X and Y directly
- W and Z were not measured directly, but we calculated their values according to the values of X and Y

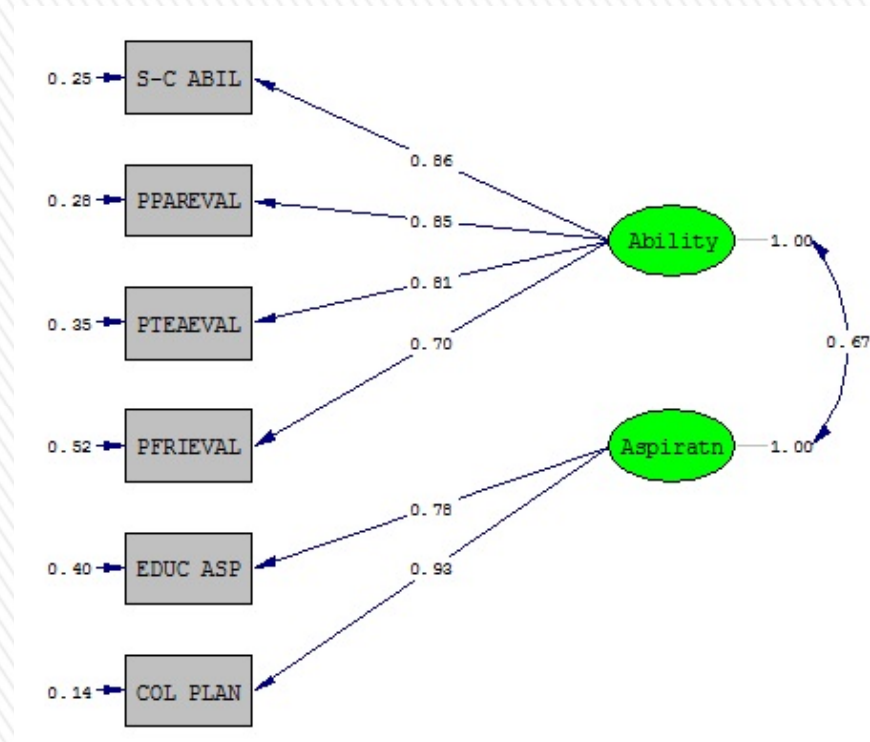


Measured and latent variables

- In our exploratory factor analysis we obtained the same number of latent variables (factors) as there were original measured variables
- We also considered how much each measured variables was related to every latent variable
- Structural Equation Modelling considers more complex relationships between these measured and latent variables

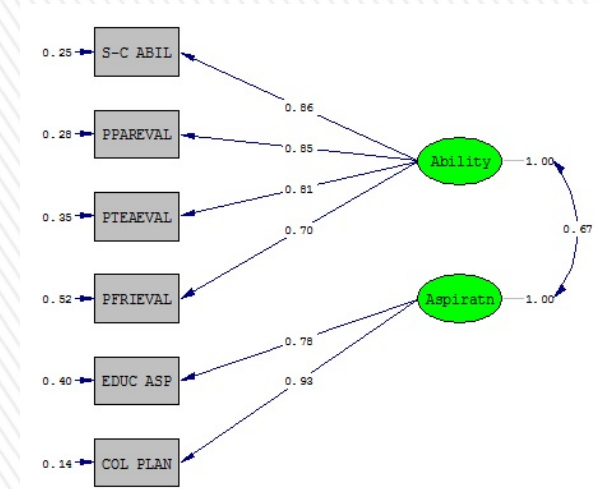


Structural Equation Modelling



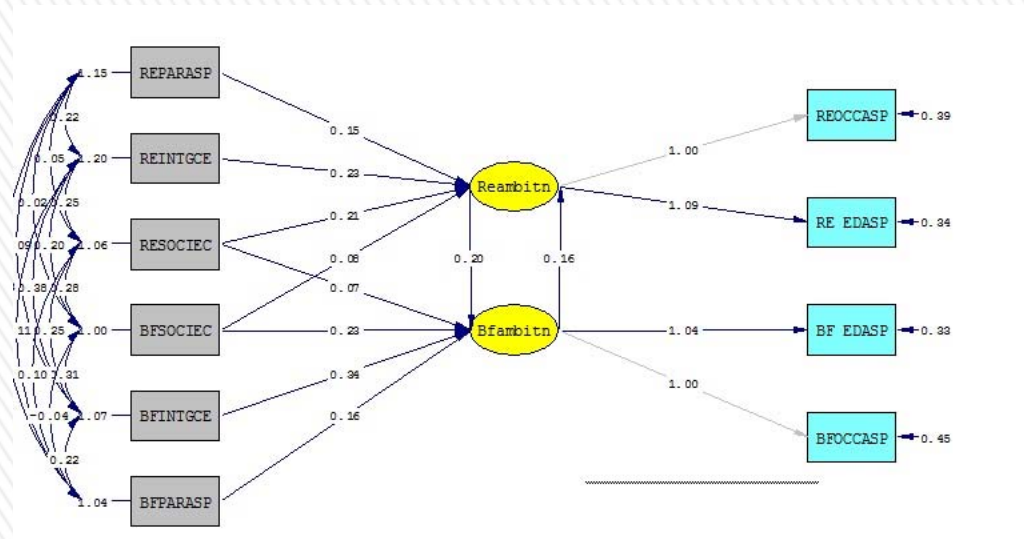
Structural Equation Modelling

- In our SEM we have:
 - a box representing each variable that we have measured directly
 - An ellipse representing each latent variable
 - A link for each relationship that we want to model (we might propose that some latent variables are only related to a particular subset of measured variables)



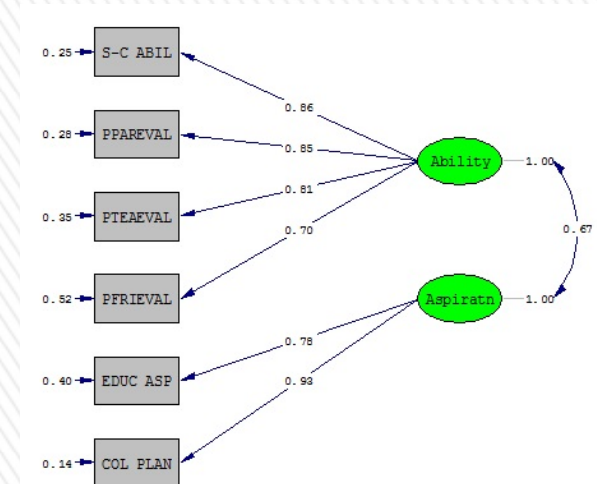
Structural Equation Modelling

- In an SEM we are free to choose any layout of relationships between measured and latent variables, and can obtain some very complex models as a result



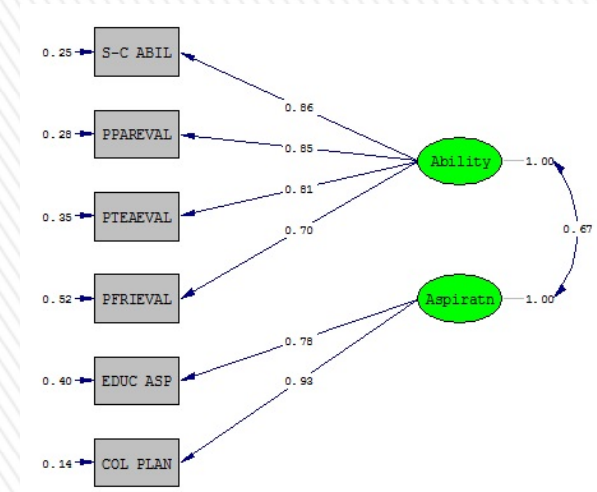
Confirmatory Factor Analysis

- A Confirmatory Factor Analysis is a particular kind of SEM
- In our original Exploratory Factor Analysis modelled a relationship between every measured variable and every latent variable (where the results indicated the strength of the relationship between each measured and each latent variable)



Confirmatory Factor Analysis

- In a Confirmatory Factor Analysis we will propose which measured variables are related to which latent variables
- We then obtain a measure of model fit (how well the model describes the data) as a result



Likert scales

- Returning to the DASS21 we remember that each question is asked on a sliding scale
- I found it hard to wind down – never, sometimes, often, almost always



Likert scales

- In reality these underlying constructs are not categorical by nature
 - If a person is depressed for a certain number of days a month, then at what threshold should a person go from often depressed to almost always depressed?
 - If I consider all of the participants that are often depressed will I still find that some participants are more depressed than others?
 - If my underlying constructs are not categorical by nature then can I ever extract these continuous variables from my dataset?



Rasch Analysis

- Rasch Analysis is used to extract these underlying continuous variables from my dataset
- Rasch uses the original categorical values (eg. for depression – never, sometimes, often, almost always)



Rasch Analysis

- Rasch analysis also uses other variables in the dataset that are correlated with the variable of interest
- For example, we might find that an older person is more likely to be depressed than a younger person

Rasch Analysis

- In this case the Rasch model would predict that an older person who says that they are “often” depressed, would be more depressed than a younger person who says that they are “often” depressed
- Hence through the use of a group of related variables, Rasch analysis is able to convert a measured categorical variable into a predicted underlying variable



Validity and reliability

- Once we obtain a given measure (eg. through methods such as a Rasch analysis) then we need to consider whether that measure is valid and reliability
- For each measure we have the measured value (DASS21) and the underlying construct (depression)
- A measure is valid if there is a strong correlation between the measured value and the underlying construct



Validity and reliability

- A measure is reliable if when we repeat the measure (possibly under slightly different conditions) then we will obtain the same value
- Construct Validity - a panel of "experts" are used to assess whether a measure truly does reflect the underlying construct
- Criterion-Related Validity - a measure correlates with another variable (as predicted)



Validity and reliability

- Formative Validity - used to assess how useful a measure would be to provide information on an outcome of interest
- Sampling validity - assesses whether a measure provides an adequate coverage of the various topics involved in a particular construct (eg. depression)



Validity and reliability

- Formative Validity - used to assess how useful a measure would be to provide information on an outcome of interest
- Sampling validity - assesses whether a measure provides an adequate coverage of the various topics involved in a particular construct (eg. depression)



Validity and reliability

- Test-retest reliability - do we obtain the same results if we administer a test on two different occasions
- Parallel forms reliability - do we obtain the same results from two different versions of a test (DASS21 and DASS42)



Validity and reliability

- Inter-rater reliability - do we obtain the same results if two different interviewers administer the test
- Internal consistency - does each individual measure within a test measure the same thing



Is my measure reliable and valid?

- Now that we have started exploring what it means for a measure to be reliable and valid, we can understand the types of biases that can creep into our study if our measure are not reliable and valid
- As we understand the potential validity of a scale we need to keep in mind that a scale is not just “valid” but is “valid for a particular population at a particular point in time”.



Valid for a particular population

- In many cases scales like the DASS21 are shown to be valid, but only for adults within an American (or English or Australia) population
- We need to ask ourselves whether the scale is still valid for children or for different populations (Asian or Indigenous Australian populations)
- Does depression have a different meaning for an Asian compared to an Australian population?



Valid for a particular population

- Invariably there will be times when we will need to adapt an existing scale to a new population
- We need to trade off the potential weakness in our study from using a scale which might not be reliable or valid, versus the cost and time required to demonstrate reliability and validity.



Valid for a particular population

- If we want to demonstrate reliability and validity then we will consider the types of reliability and validity discussed on the previous slides (eg. test-retest reliability and construct validity).



Valid for a particular population

- We might also want to consider whether our statistical methods (Factor Analysis and Structural Equation Modelling) produce similar results).
- Is there a strong relationship between a given item and an underlying construct for one population, but not for another?



Conclusion

- Designing good questions for a scale
- What is the difference between exploratory and confirmatory factor analysis?
- What is Structural Equation Modelling?
- What is Rasch analysis?
- What does reliability and validity mean?
- Adapting an existing scale to a new construct or a new population

