Attribute Identification and Validation

NCME 2009
Workshop

Introduction

- Basic discussion of what is a reasonable attribute.
- Purpose matters
- The definition of an attribute
- Validation of the attributes
- Comment on what happens if we already have the test?

Definitions

- Attribute: a *categorical latent variable* representing the diagnostic status of a person.
  - *Categorical*: can be one of a finite number of discrete levels (usually two)
  - *Latent*: not directly observable
  - *Variable*: status changes from person to person

- Synonyms for attributes:
  - Skills
  - Factors
  - Diagnoses

Types of Attributes

Table 4.1

Exemplary Attribute Definitions from Different Domains

Construct: Number Subtraction Domain: Mathematics Source: de la Torre & Douglas (2004)	
- convert a whole number to a fraction - separate a whole number from a fraction - simplify before subtracting - find a common denominator	- borrow from whole number part - column borrow to subtract - subtract numerators - simplify answer
Construct: Reading Comprehension Domain: English Language Learning Source: Buck, Tatsuoka, & Kostin (1997)	
- synthesize scattered information - recognize relevant information - know low-frequency vocabulary - identify the gist of a passage	- apply relevant background knowledge - hold relevant information in WM - use a word-matching strategy - compare two plausible options and choose
Construct: Figural Analysis Domain: Architecture Source: Katz, Martinez, Sheehan, & Tatsuoka (1998)	
- move or rotate objects - read and translate information - activate prior knowledge - identify distracting information	- identify environmental characteristics - process a complex diagram - understand structural technology - apply a learned procedure

Grainsizes of Attributes

- The degree of definitional specificity of an attribute is often referred to as the *definitional grainsize*.
- The grainsize is driven by the level of specificity with which one would like to make statements about respondents.
- The grainsize of an attribute is the resolution with which an investigator dissects a cognitive response process and describes its constituent components.

Practical Issues with Grainsizes

- It is possible to decompose individual attributes for more complex tasks further
  - That would increase the number of attributes.
- As the number of attributes increases, the number of latent variables in a DCM increases.
  - Attribute profiles and item parameters may become impossible to estimate statistically.
- It is important to fix the number of attributes to a statistically manageable number for a given diagnostic assessment length and respondent sample size.

How to define attributes?

- Think from the beginning what you want to say about who and why
- Think about what evidence is needed to say this and how to get it
- What you plan on doing with this when you get it

Purpose

- Here we will discuss two goals for defining attributes
  - A cognitive model has been defined in which the attributes represent the underlying processes needed to answer each question
  - The test is being used to assess ones “ability” (mastery or nonmastery) of a set of previously defined goals or objectives

Purpose

- We note that the purpose matters
- Here we first focus on the situation where the true underlying process and thus the needed attributes to complete any item are defined
- This method is heavily rooted in the Cognitive Design System

The Cognitive Design System

- Strongly associated with work done by Susan Embretson
- Grounded in Applied Cognitive Psychology
- Revolves around the idea that the basic cognitive abilities of an event can be identified and manipulated
  - In doing so items can be automatically constructed with “known” characteristics

Steps of the Cognitive Design System

1. Specifying the goals of measurement
2. Identifying relevant features in the task domain
3. Developing a cognitive model of task performance
4. Generating items according to the cognitive model
5. Evaluating the cognitive model empirically via administered tasks
6. Banking the items by cognitive complexity
7. Validate the model by checking for nomothetic span (criteria validity)

Example

- Measure Fluid Intelligence
- Solve new tasks or recognize new patterns in visual patterns (Progressive matrices)

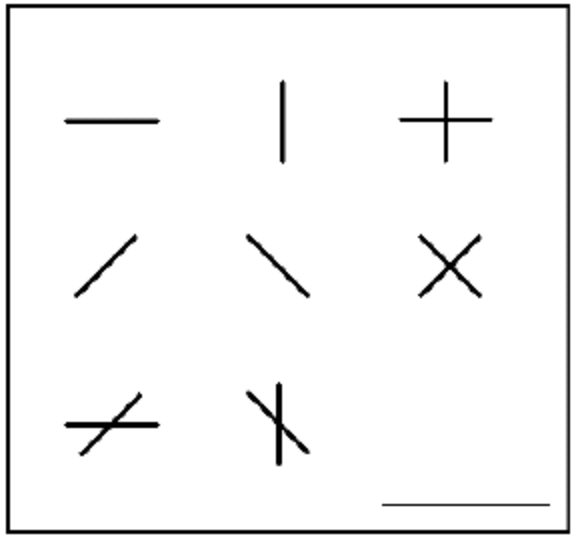

Figure 2.2 A basic task in an abstract reasoning assessment for fluid intelligence.

Underlying Process

- To complete this method there must be a general agreement of the process/cognition used to answer an item or it must be based on your own theory
- In any event to develop a general agreement we recommend using a panel of experts on the general construct of interest
- Given this panel a Dephi Method or Concept Mapping can be used to indentify the underlying process

Delphi Method

- Pose a question to the Panel (each person is to answer the question independently)
- Summarize responses and return to the panel highlighting difference and similarities
- Ask panel to respond
- Repeat process as needed

Concept Mapping

- Using a panel hold a “Brain Storming” Session
  - All ideas or statements that are mentioned are recorded
- Eliminate all redundant concepts are eliminated
- Ask the panel and/or additional experts to sort concepts based on similarity
- These similarities are analyzed to obtain a set of underlying attributes

Example

- Cognitive model to tasks
  - Identity (i.e., recognizing each shape),
  - Addition (i.e. adding two shapes to get the third)
  - Progression (i.e., recognizing how shapes change from column one to column two to column three),
  - Distribution of three (i.e., recognizing that shapes reappear in threes across the rows of the figure), and
  - Distribution of two (i.e., recognizing that two similar shapes appear in
- Use these to create items that vary in complexity and difficulty and assess

Using Defined Attributes

- Next we consider an approach that could be used if the attributes are already defined
- A situation similar to this may occur if you are interested in developing a benchmark test intended to assess mastery or nonmastery of a set of state defined objectives or goals
- An approach that would work well in this case is rooted in Evidence-Centered Design

Evidence-Centered Design

- Strongly associated with work done at ETS including Bob Mislevy
- Developed to structure the thinking for those creating an assessment
  - Diagnostic Narrative
- Design tasks that maximally elicit certain behaviors

Assessment Narrative

- Use *Evidentiary Arguments*
  - *Made of a claim and warrants (i.e., the data)*
    - *If an examinee is a master of a particular attribute they are expected to be able to complete this particular activity*
    - Jamie has most likely mastered basic addition (*claim), because she has answered correctly a mathematical problem about adding up prices in a supermarket (data)*

Assessment Narrative

- *In addition Backing and Refusals (a justification for the claim and warrant and explanation why an alternative is not true)*
  - It is most likely that she did this because she applied all of the individual addition steps correctly (*backing) and the task was designed to force her to do that (backing).*
  - She may have used her background knowledge to estimate the final price of her shopping cart (*alternative explanation), but that is unlikely given that the final price is exactly correct (refusal)*

ECD Components

When completed ECD should have a detailed set of definitions of the attributes that contain:

1. The student models, which formalize the postulated proficiency structures for different tasks (the learning theory and objectives)
2. The task models, which formalize which aspects of task performance are coded in what manner (e.g. the test)
3. The evidence models, which are the psychometric models linking those two elements (evaluation)
4. The assembly model, which formalizes how these three elements are linked in the assessment
5. The presentation model, which formalizes how the assessment tasks are being presented

Validation

- Provided that we have a set of well defined attributes of “reasonable” grain-size we must validate our definitions
- This can be done using two steps
  - Item Construction
  - Item Evaluation

Item Construction

- A set of individuals are asked to construct a set of items that assess the attributes
- In doing so, a subset of items that would be contained in the assessment are created.
- Feedback is obtained related to the definitions and their usefulness in creating the assessment.

Item Evaluation

- In addition, experts are provided with the definitions of each attribute and a set of items
- They are asked to identify which attributes each item measures
- The agreement between experts and their feed back is used to refine the definitions of the attributes.

Validation

- Once these procedures are completed, a set of well defined attributes have been developed and validated.
- Therefore, it should be possible to construct a test to assess the attributes of interest
- However, there may be times when a test has already been developed

Previously Designed Assessments

- In these circumstances the previous methods could be repeated and then matched to the Assessment.
- As an alternative, it is possible to ask experts what attributes are being measured by the test.
  - Work through the items
  - Talk out loud
- In doing so a Delphi method could be used to identify and define the attributes measure by the assessment.

Previously Designed Assessments

- We give a brief word of caution.
- From our experience, a common situation where skills are unknown is a unidimensional test.
  - One would like additional information about the examinees while also getting the unidimensional ability.

Previously Designed Assessments

- Some difficulty may arise if a test was initially developed to measure a continuous unidimensional skill and now the purpose is to determine multiple dichotomous skills.
  - The basic result will be categories that can be defined as a discrete ability scale.
- Diagnosistic models are most beneficial for tests that are not truly unidimensional.

Summary

- Although largely overlooked, attribute development and validation is critical to success of an assessment
- Great care must be used in developing a detailed definition of each attribute that can be used to both develop new items and to evaluate new items.