

## CHAPTER 7: HISTORY OF LANGUAGE TESTING

### 1. GRAMMAR-TRANSLATION

- faculty psychology and traditional grammar
- subjective tests → no item analysis, reliability, and validity → **intuitive era**
- Facts about English often weighed as heavily as skill in using the language.
- common techniques → translation, essay, dictation, précis, and open-ended answers based on reading comprehension
- the **pre-scientific** stage

### 2. DISCRETE-POINT APPROACH

- Behavioral psychology and structural linguistics
- Structural linguists → language elements could be separated (hence the name discrete-point tests) → **Skills/components model**
- Assumption → a collection of discrete points if taken together as a single score, will produce a measure of some global aspect of language ability
- The key concern of educational measurement experts (or psychometricians) → provide objective measures using various statistical techniques to assure reliability and certain kinds of validity
- Decontextualized items
- The **scientific** stage
- Criticisms:
  - Linguistically, GT theory questioned the validity of the structural view of language → the assumption that language can be divided into isolated bits could no longer be acceptable
  - Psychologically, Gestalt psychology believed total language ability is different from the sum of the discrete abilities → the contribution of every single element to the total language use and act of communication would not be clear
- Advantages:
  - desirable psychometric characteristics
  - extremely practical
  - diagnostic value

### 3. INTEGRATIVE APPROACH

- Cognitive psychology and generative grammar
- Language is a dynamic system whose various sub-skills are assessed all at the same time → **integrative/global test**
- Oller (1979) believed in a g-factor (or general language proficiency) or the idea that second language proficiency is a unified set of interacting abilities that cannot be tested separately → **unitary trait hypothesis**: an **indivisible** view of language proficiency
- A concern for *meaning* and the *total communicative effect* of discourse as in cloze test

- **Grammar of expectancy** → knowledge of language which includes knowing whether a word or utterance is likely to occur in a particular context or situation → the measuring instrument has been called **Pragmatic tests** which invoke linguistic sequences that relate to extralinguistic context in meaningful ways  
Oller claimed that the idea of making guesses in context-reduced situations provides evidence for a general language factor or the idea that second language proficiency consists of a single global trait (i.e., Unitary Trait Hypothesis).
- Limitations:
  - time required to administer and score such tests
  - reliability of cloze and dictation tests was in question
  - did not tap the ability to really use the language

#### 4. FUNCTIONAL-COMMUNICATIVE APPROACH

- Linguists → concerned with the analysis of human discourse and communicative interaction.  
Psychologists → language as a task the subjects learn to communicate, out of which syntactic structures are developed.
- The testee is expected to *perform certain functions* in order to communicate → communicative competence is defined as the knowledge about the form of language and about how to use it properly in context.
- **Needs analysis** → the relevant functions and the domain and conditions under which the functions have to be performed are identified → English for specific purposes (ESP)
- Measure different language skills in communicative tests based on a view of language referred to as the **Divisibility hypothesis** → obtain different profiles of a learner's performance in the language, i.e. profile-reporting.
- **Authentic materials**
- Concept of qualitative modes of assessment in preference to quantitative ones → **language band system** is used to show the learner's levels of performance in the different skills → detailed statements of each performance level serve to increase the reliability of the scoring.

### Functional Testing

#### 1. DEVELOPING TEST STEM

##### 1.1. Pretesting with Native Speakers

Elicit socially appropriate and linguistically accurate responses for items

##### 1.2. Pretesting with Non-native Speakers

# Elicit the possible inappropriate answers for items

# Responses given by native and non-native speakers are categorized into four classes:

- *Functional responses* → socially appropriate and linguistically accurate responses
- *Linguistic responses* → socially inappropriate and linguistically accurate responses
- *Social responses* → social appropriateness and linguistically deviant responses

- *Distractors* → both socially and linguistically deviant from the native speakers' norm

### 1.3. Pretesting with Native and Non-native Speakers

Main purposes: (a) to assure that NSs would select functional responses and (b) to make sure that NNSs should understand the directions and the purpose of the test

## 2. SCORING SYSTEM

A weighted scoring system is usually used in functional tests:

- functional responses receive two points;
- either linguistic responses or social responses receive one point;
- distractors receive no points.

☛ **Note:** The Functional test incorporates contextualization, the most praised characteristic of integrative test, and item independency, the obvious advantage of discrete-point tests.