



CLASSROOM TESTING

Introduction

- Would you describe the tests used in your classes as:
 - Discourse-driven?
 - Functional?
 - Representative of how and what you teach?

Introduction

- Why are the answers to these questions almost always “no”?
 - A preference by students or teachers for discrete-point testing
 - Difficult and time consuming to design and grade discourse-based and functional tests
 - Lack of widely accepted, practical model of communicative testing
- (Magnan, 1991)

Our Goal

- To find ways in which tests and other types of assessment can be designed to reflect course goals most effectively.
- How well are we doing?

Language Testing: Some General Concepts

- It is important to consider the purpose of tests.
- What is the primary function of a test?
 - Administrative purposes (placement, general proficiency, certification, promotion)
 - Instructional purposes (classroom tests)
 - Research purposes (evaluation of programs, experimentation)

Test Designs

- Direct tests – incorporate contexts, problems, solution strategies that students would use in real life (authentic assessment, alternative assessment)
- Indirect tests – represent competence by extracting knowledge and skills out of their real life contexts (pen & paper)
(Liskin-Gasparro, 1996)

Test Design

- Discrete point – test one and only one point at a time (isolated grammar and vocabulary, sociocultural knowledge)
- Integrative – testing more than one point at a time (giving a talk, cloze passage)
(Cohen, 1994)

Test Types

- Achievement – classroom tests evaluate students' acquisition of course content (only what is covered in class)
- Proficiency – tests to measure an individual's general competence in a L2; are independent of a course of study (ACTFL OPI Interview)
- ▶ Which do we use in our classes?

Traditional Achievement Tests

- Decontextualized
- Discrete-point
- Fill in blanks, translation
- Concern only with accuracy
- No concern for sociolinguistic considerations or discourse competence
- Individual test sentences do not have meaningful content
- Do not represent real language use

BUT, we aim for language testing *in context*

- Tests should reflect both accuracy and language use in context
- Contextual testing began in 1975 (Bondaruk, Child, Tetrault) – embedding test items in discourse length contexts as opposed to single phrase. They use cloze procedures with authentic texts.
- Use contextualization -- should apply to both testing and textbooks (Walz, 1989)

Language testing in context

- Test design should ensure that students need to understand the meaning being conveyed to accomplish the task.

Types of Evaluation

- Formative – enables teachers to alter instruction in progress to better address the needs of learners (more limited in scope)
- Summative – characterize the language abilities of individuals at a given point in time without regard to specific course of instruction (proficiency tests)

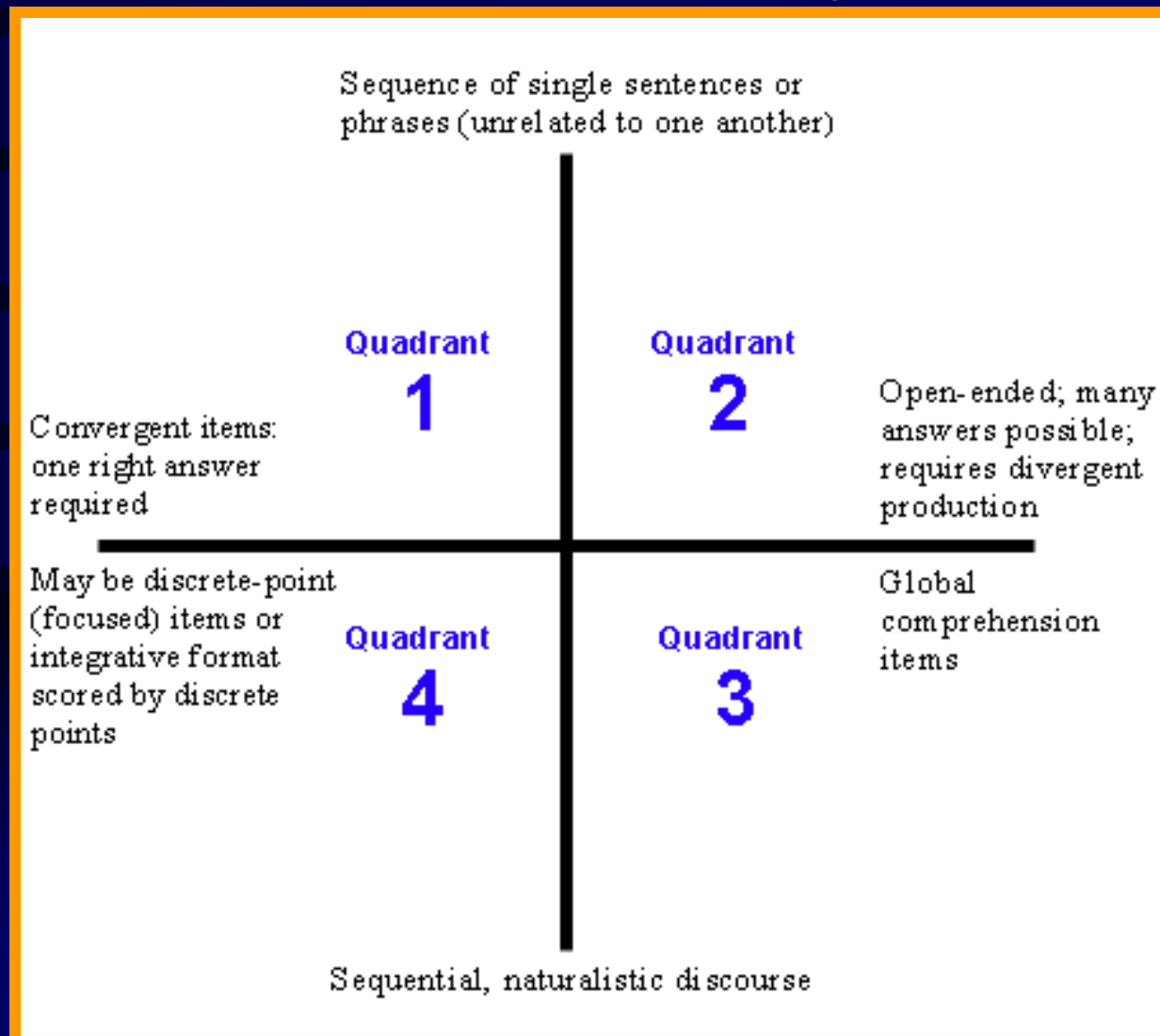
Making the case for Hybrid Classroom Tests

- Construct tests to integrate lexical, grammatical, sociolinguistic and discourse features so that they operate in naturalistic discourse contexts
- Test in situational formats (since language occurs in discourse context)
- Students are more motivated to learn what is relevant to their communicative needs

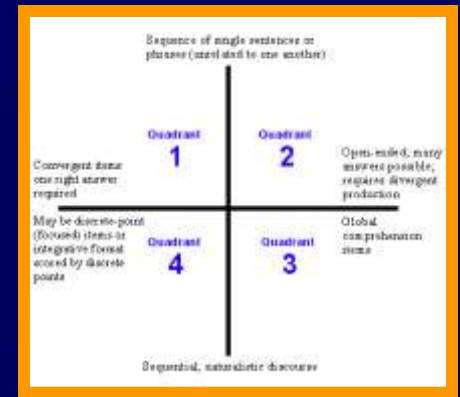
Hybrid Classroom Tests

- Combine discrete point, free response
- Aim for a blend of strategies (use a variety of formats to test students work)
- Incorporate alternative assessment such as portfolios, samples of oral performance, exposés

Characteristics of Test Items and Item Types

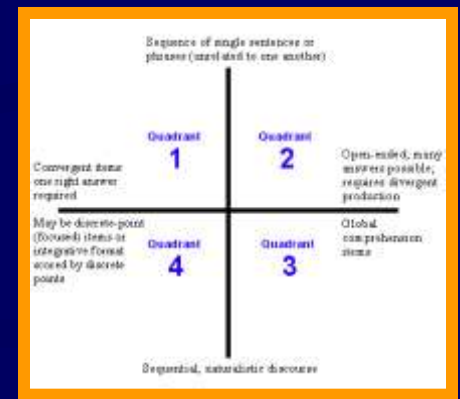


Characteristics of Test Items and Item Types



- Vertical axis: naturalness or authenticity of language used on test (from drill like – upper portion – to natural discourse – lower portion)
- Horizontal axis: left – most convergent or discrete point; right – more open ended with focus on communication

Characteristics of Test Items and Item Types



- Tests should include items from various points on horizontal axis to blend communicative and linguistic concerns
- All items, however, should fall in lower ranges of vertical axis so that natural language is represented
- Contexts should be chosen to reflect themes of lessons or units of study, as well as functional and communicative purposes

How well do we measure up?

- Think of the last test you administered in the class you teach (or a test you have taken?):
 - Were exercises contextualized?
 - How many discrete-point items were there?
 - What types of exercises were provided? (fill in blank, translation, cloze, etc.)
 - Was there a concern for more than just accuracy?
 - Was there a concern for sociolinguistic considerations or discourse competence?
 - Did the individual test sentences have meaningful content?
 - Was real language use present?

Was it traditional or proficiency based?

Sample formats

Listening Comprehension Formats

- Listening for specific grammatical or lexical features
 - Read a passage and students listen for specific features
 - Conversations work well
 - Passage should be in context, as opposed to reading individual isolated sentences
 - Ex: recorded message for train schedules, check off items to buy, things that students will do or have already done, students list information about apartments

Listening Comprehension Formats

- Listening and Following a Map or Diagram or Choosing or Completing a Picture
 - Synthesize cues
 - Follow directions with a penciled line on a map
 - Fill in names of stores, buildings
 - Enter a set of names in a diagram of family tree
 - Complete a room plan
 - Select picture most accurately described

Listening Comprehension Formats

- Comprehension Questions
 - True/false, multiple choice, completion questions
 - If testing language comprehension, answers should be elicited in students' native language
 - Keep in mind purposes for listening to type of material chosen in real life
 - Ex: news broadcast – listen for main ideas

Listening Comprehension Formats

- Writing a summary of a message
 - Summary of facts and/or inferences
 - Also known as “immediate recall protocols”
 - Grade for main ideas and details

Listening Comprehension Formats

- Global Classification and Gisting
 - Report gist of listening passage or classify it globally
 - Ex: Students give a title to a passage

Reading Comprehension Formats

- Reading for Precise Information
- Reading for Sociocultural/Pragmatic Cues to Meaning
- Reading and Choosing the Best Paraphrase

Writing and Mixed Skills Formats

- Require written responses in the TL
- Can range from very controlled to open-ended exercises
- Ideas presented may be supplemented with ideas for writing

- Sentence Cues: telegraphic sentences, somehow thematically linked
 - Advantage: help lower proficiency students
 - Disadvantage: difficult to contextualize, unrealistic
- Contextualized Partial Translation: elements of discourse deleted and must be restored using partial translation cues
- Cloze Exercises: elements of discourse deleted, options for filling in provided in a box
- Discourse Transformation: passage must be somehow changed – i.e., replace nouns with pronouns, change present to past, etc.

- Partial Conversations: students see half of a conversation, must provide other half; more creative and open ended
- Translation Cues: provide cues for students to formulate specific responses, questions, etc.
- Open-Ended Completions: students supply logical ending for set of partial statements; can be related to reading passage; can be personalized
- Responding to a TL text: students write response to reading in TL; combines both skills in contextualized format.

Contextualized Paper & Pencil Tests

- We would all agree that a *contextualized* test is the best kind, but how do we create them?
- With a little practice, creating contextualized tests is actually easier, since they integrate so many linguistic and lexical features in a small number of well-chosen passages.

Guidelines for creating contextualized tests

- 1. Take an inventory of the material
 - List major grammatical, lexical, discourse, cultural features you'd like to cover
- 2. Decide which skill areas will be emphasized.
- 3. Divide the content among the testing types you have chosen.
- 4. Create a simple story, dialogue or paragraph that includes the features you want.
- Keep in mind that the test should not differ drastically from activities students have done in class.

The next step: Evaluating the draft

- How does the test as a whole sample students' use of language?
- Is evaluation/credit awarded fairly?
- What is the fairest way to grade a test?
- How can we ensure grading uniformity across multi-section courses?
- Don't be afraid to rewrite, redo and edit!

Testing Oral Skills

- Oral exams can be among the most challenging to create, schedule, administer and grade.
- Yet oral proficiency, we have all agreed, is one of the most important goals of our classrooms.
- So how can we test oral abilities?

The ACTFL OPI

- Only certified and trained testers/raters can administer real OPIs, but to get an idea of how they work...
- Face-to-face conversation of 10-25 minutes
- Complex and purposeful structure
- Four phases, all of which fulfill psychological, linguistic and evaluative functions.
 - Phase 1: Warm-Up (~ 3 min.)
 - Phase 2: Level Check (repeated)
 - Phase 3: Probes (repeated)
 - Phase 4: Wind Down (~ 2 – 5 min.)

Adaptations for classroom testing

- Conversation Cards and Interviews
- Individual Interviews
- Paired Interviews
- Description of picture, drawing, etc.
- SCORING:
 - Taped tests?
 - Evaluation sheet should be clear, and available to students beforehand

Recent Innovations in Testing

- Computer-Adaptive Testing
- Alternative Assessment Measures
 - Portfolios
 - Oral proficiency measures
 - Performance-based tasks

What do you think?

- What are the major advantages and disadvantages to:
 - Standard written testing techniques?
 - Integrative contextualized (mixed skills) tests?
 - Computers in testing?
 - Alternative proficiency measures?