1. INTRODUCTION

1.1 Popular testing procedures

- traditional fill-in-the-blank grammar tests
- multiple-choice and extended answer, comprehension questions on a script or a text
-aural comprehension drills of various kinds
- structured and open-ended writing tasks, usually in response to a prompt
- structured or improvised oral interviews

The routine use of these methods of assessment of Spanish language ability does not necessarily make them reliable and valid.

1.2 Models of Communicative competence in SLA

Swain (1980, Canale, 1983):
1. grammatical competence, morphology, syntax, lexicon, phonology,
2. sociolinguistic competence (appropriate use of language),
3. strategy competence (cognitive and metacognitive strategies used when communication breaks down)
4. pragmatic competence (oral and non-verbal coping mechanisms used when communication breaks down)

Buchman (1989) revised model of communicative ability
1. organizational competence
   a. grammatical competence
   b. discourse or textual competence
2. pragmatic competence
   a. sociolinguistic ability
   b. illocutionary competence
3. strategic competence

He and Young (1998) added competence in (at least) the following five interactional features:
1. Knowledge of rhetorical scripts,
2. contextually-relevant lexicon and syntax,
3. strategies for managing turns,
4. management of conversation topics and
5. the means to signal boundaries in a conversation

2. METHODS TO ASSESS CLASSROOM LEARNING

2.1 Discrete-item and integrative tests

Here are a few examples of integrative testing formats:
- Choice
- Dictation
- Summary

2.2 Using tasks as tests

Hughes (2003) proposes the use of four writing tasks centered around the theme of work at a summer camp for children:

2.3 Real-life tests
1. face-to-face ACTFL-OPI
2. tape-mediated SOPI
3. Mixed form: students interacting with each other with prompts from a computer or a tape

Advantages of semi-direct tests:
1. It allows for a uniformity of elicitation procedures, which helps in producing validity.
2. It is economical to administer since there is no computer test administration to interfere with each respondent.
3. It eliminates the interview effect which can play a role in oral interviews.

2.4 Complementary measures to assess second language ability

Method effects result from differences in:
1. discourse tasks (e.g., reporting vs. interviewing),
2. elicitation methods (e.g., personal vs. machine),
3. genre (e.g., narrative vs. expository texts),
4. item types (e.g., multiple-choice vs. open-ended tests), and
5. test consequences (e.g., declared vs. undeclared purposes of tests)

Solution: use of a battery of alternative assessment instruments in complementary ways
2.5 Assessing the usefulness and relevance of a test:

Six qualities that will determine the value of the instrument:
1. **Reliability**: the consistency of measurement,
2. **Content validity**: an indicator of the ability we want to measure,
3. **Authenticity**: the correspondence between the characteristics of the test task and the features of the real task,
4. **Language varieties**: the interaction between the test-taker and the task - including language ability, topic knowledge, and the affective situation,
5. **Impact**: society and the individuals, and
6. **Practicality**: the demands of task specifications can be met with existing resources.

- Large-scale testing: reliability and validity are likely to be crucial.
- Most types of classroom testing, authenticity, interactivity, and impact are the likely factors to be most relevant.

2.6 Washback effect

The link between curriculum and assessment of the objectives pursued in a curriculum needs to be bidirectional.

3. TESTING PRAGMATIC - CULTURAL KNOWLEDGE

3.1 Can we test for knowledge of Spanish pragmatics?

 Etherod & Marquez Reiter (2003: 335) analyzed the phrasing of the phrase, "How is the weather today?" in Spanish and English.

I was just wondering if you have the book, and if I could borrow it.

"¿Cómo va el clima hoy?" Can you lend me the book?

¿No has pensado en leer este libro?

Haven't you thought about reading this book?

- Raffi (1988): The OPI did not properly or thoroughly evaluate sociocultural and sociolinguistic abilities. Expand the use of the roleplay task.

- Teaching and testing pragmatics sends an important signal to the students that the behavior being assessed is worthy of such assessment.

3.2 Can we test cultural knowledge in Spanish?

North (2000: 95): "Intercultural skills are an aspect of socio-cultural competence not found in any of the scales analyzed.

Objective of teaching and testing culture:

1. Learning a second language cannot be simplified reduced to becoming a monolingual speaker of that language.
2. The assessment of intercultural skills is multifaceted.
3. There are right and wrong answers when it comes to testing culture.

North (2000: 13): "scales of proficiency "have the potential to exert a positive influence on the orientation, organization and reporting of language learning." North sees the potential for scales in that they provide learners with:
1) explicit goals and descriptions of them,
2) coherent internal links for curriculum development and testing,
3) behavioral evidence of progress,
4) a means for increasing the reliability of subjective ratings,
5) a common metric for comparisons among different populations of learners.

3.3 Developmental nature of institutionalized rating criteria

1. Nation of implicit degree of complexity of tasks, general goals or focus.
2. No organized way to data that can be used to develop proficiency scales, and to identify stages that one may achieve in formal registers. By that step, we will be able to discuss concrete and factual topics, before they manage more abstract topics or from familiar to unfamiliar.
3. Levels may not be hierarchical, but rather complementary as they represent different registers of the language.

4. PRACTICAL RECOMMENDATIONS

1. Give primary importance to the context of learning.
2. Identify, describe, and operationalize learning goals as testing objectives.
3. Test the course objectives.
4. Obtain a robust sampling of selected course objectives.