

Oral tests for children learning a foreign language

Rocío Domínguez
Carnegie Mellon University

Abstract

During the past two decades the impact of the communicative competence approach in foreign language instruction has increased the demand for oral assessments. Approximately 70% of the children in Foreign Language in the Elementary School programs in the USA are involved in some kind of oral test. In this article, a rationale for the assessment of oral skills for early language learners in Foreign Language in the Elementary School programs is provided. The factors extensively discussed here are discourse domains of oral tests, child's level of cognitive development and type of test. An special emphasis is given on the interrelationship of these factors as well as their implications for test validity and reliability.

1. Introduction

In the past decade, the incidence of foreign language instruction in elementary schools in the U.S.A. has increased by nearly 10% from 22% in 1987 to 31% in 1997 (Rhodes and Branaman, 1999). This increase has raised questions regarding the appropriate method of foreign language (FL) instruction and as a consequence the adequacy of assessment tools for young foreign language learners (five to ten years old). In a survey given to 2932 Foreign Language in the Elementary School (FLES) programs in the U.S.A. in 1997, Rhodes and Branaman (1999) report that approximately 70% of the children from those elementary schools are involved in some kind of oral assessment including presentations, demonstrations, authentic activities, and oral interviews. The language skills involved in those tasks are listening and speaking. For young FL learners, assessing writing skills in the foreign language is out of question since FL writing is introduced in later grades (5 to 6) in the curricula. The high percentage on oral assessment is not surprising if one takes into consideration the impact of the communicative competence approach in FL instruction for the past two decades (Thompson, 2001).

Furthermore, research in second language acquisition has shown that children and adults acquire a foreign language *differently* (McLaughlin 1978). It is not appropriate then to use teaching methodology nor methods of assessment for adults to measure FL proficiency of young learners. Rather the assessment tools should be developed for the test takers in an appropriate manner. Carpenter, Fujii and Kataoka (1995) report that young Japanese learners behave very differently from adults when being tested orally (see below). With such a scenario, the demand for research in the assessment of speaking skills for children is sustained. It is then comprehensible the importance of answering questions related to the appropriate assessment of child oral skills in FL learning. For instance, it urges us to define the domains of speaking in the classroom environment within the current theory of communicative competence. What kind of 'speaking' are young children are involved in? Does speaking in the classroom resemble real-life conversations? If not, what are the features of speaking in the classroom for young foreign language learners? How much time do students speak during the language lesson? How many opportunities are students given to talk the target language? (Swain; 1988, 1996a, 1996b). It is not the purpose of this paper to question practices in FL instruction, rather to argue of the necessity of making oral tests similar to classroom practices so that test validity not be endangered.

Throughout this paper, a rationale for the assessment of oral skills for early language learners in FLES programs is offered. Factors involved in testing children orally are discussed. A final comment is made on a political issue regarding oral assessment: the importance of showing children progress in acquiring the FL might be related to the necessity of financial support for continuing language programs. Thus, it is likely that educational agents, who construct oral tests, may want to maximize the amount of target language elicited in the test in order to demonstrate students' progress. In turn, this practice may conflict with the validity of the test since children have not been exposed to a method of instruction that prioritizes student talking in the classroom. There might be then two forces in tension with regard to child oral assessment. One force might call for manipulating the test with political purposes, while a second force might call for constructing the assessment according to what and how children have learned. I am more inclined to the second force simply because testing children in what they have not learned is not only unethical but may have catastrophic consequences.

2. Factors involved in testing foreign language for young children

Assessing young FL learners orally is not an easy task. Children are still increasing their knowledge in their native language as a reflection of their expanding knowledge about the outside world. When attending school, children open the door of a new world of language: they have access to new registers (e.g. written and oral standardized language), which entails new vocabulary, different styles, diverse grammatical structures, as well as various contexts. Furthermore, children change rapidly both physically and mentally. All these changes are reflected in the way they use language too. How can all these variables be captured when testing children? This work explores those variables and how they are interconnected. In order to understand one of them, one must understand the others as well. The factors extensively discussed in this paper are *domains of oral tests*, *child's level of cognitive development*, *issues of the test*, and *validity*. Some issues on *reliability* are discussed briefly. All these factors are shown in Figure 1.

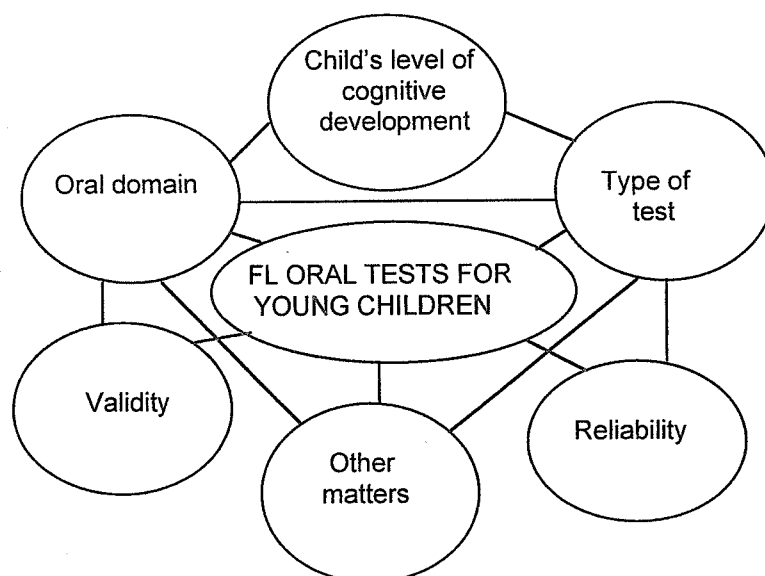


Figure 1. Factors involved in FL oral tests for young children

2.1 The oral domain

In this section, the language ability used in language testing for young FL learners is discussed following Bachman's (1990) model of communicative language ability in language testing. This model responds to the necessity of creating a theoretical basis for language testing consistent with the theory of communicative competence (Hymes, 1972; Canale and Swain, 1980; Canale, 1983). This necessity is the result of the impact of the communicative competence theory on teaching methodologies mentioned in the introduction. A brief presentation of Bachman's model is provided next.

Bachman states that in his model the ability to use language communicatively is defined as the knowledge or competence of the language, and the capacity for implementing, or using this competence. This model breaks with traditional models of language in which language skills (listening, speaking, reading and writing) were distinguished from language components of knowledge (grammar, vocabulary and phonology) (Lado, 1961; Carroll, 1968). Such models not only failed to provide accurate definitions of the so-called language skills, they also did not indicate how skills and components are related. Theoretical inconsistencies in these models were evident when attempting to operationalize the skills. For example, how do we define reading? What are we measuring exactly when testing reading?

Instead, Bachman proposes that communicative language ability can be described as consisting of three components: language competence, strategic competence, and psycho-physiological mechanisms. Language competence in turn consists of other components such as organizational competence and pragmatic competence. Organizational competence comprises abilities concerning the production of "grammatically correct sentences, comprehending their propositional content, and ordering them to form texts." (Bachman, 1990, p. 87.) Pragmatic competence comprises abilities that concern the relationship among signs and their referents. Strategic competence in turn is a complex component that mediates between the language competence and the psycho-physiological mechanisms at any given situation. This component refers to how individuals display their language competence. The psycho-physiological mechanisms are essentially neurological and physiological processes such as the visual and auditory channel, the receptive and productive mode. Is not the purpose of this paper to discuss these concepts in depth, rather to make connections with language testing for young learners. The model of communicative language ability serves here as a reference for framing language

testing for FL young learners. Next, the notion of language ability for children learning a FL is narrowed down, in order to establish possible language ability' domains for oral tests. Currently, language theories that emphasize communicative approaches of language agree on that language cannot be separated from the context of its use. According to Bachman (1990) *language use* "is the creation or interpretation of intended meanings in discourse by an individual, or as the dynamic and interactive negotiation of intended meanings between two or more individuals in a particular situation" (p. 61). It is then indispensable to conduct studies on child interaction in oral tests. In order to understand what can we expect from children in terms of speaking we have to understand speaking in a way that is appropriate for children in a testing situation.

According to Bachman, there are two major issues for defining the domain or domains of the language ability to be tested. First, this ability needs to be sufficiently delimited in order to distinguish it from other individual characteristics that may affect test performance. Second, the language ability needs to be appropriately defined for each particular testing situation. Young learners of foreign languages do not use the FL at all outside the language class, thus, it would be more useful in this case to base the test tasks exclusively upon language use in the classroom. The domain or domains of speaking are then delimited by classroom dialect. Speaking in the classroom may be slightly different from real-life conversations. In addition, the way or ways speaking in the FL is promoted in the classroom are mediated by a teaching methodology that in turn is related to instructional objectives. Thus, it is important to define the domain or domains accordingly to what is feasible from the student perspective as it has been observed in class (Messick, 1996). However, the definition of domain is not as clearly cut as one would desire. Douglas and Selinker (1985) suggest the following definition of domain:

"A discourse domain is a personally, and internally created 'slice' of one's life that has importance and over which the learner exercises content-control. Importance is empirically shown by the fact that in interaction one repeatedly talks (or write) about the area in question. Discourse domains are primarily dynamic and changing, and may become permanent parts of a learner's cognitive system. Some domains may be created temporarily for particular important purposes. The concept also has a discontinuous aspect to it in that a domain can be taken up, dropped, left dormant and revived. Such domains are usually thus not fixed for life but may change with one's life experience – and often do." (p. 206)

The first feature of this definition of domain that calls our attention is its individual nature. It is important to notice that this feature will cause the definition of domain to be impossible to capture fully from the outside, because we cannot look inside the brain of the individual. Another reflection is that the internal creation of the domain cannot be possible without its interaction with the context of the situation. The individual nature of the domain also goes hand in hand with being dynamic and important to the individual. Both characteristics and the domain being temporary apply for the testing situation allowing the testee to make sense of the questions of the test. Douglas and Selinker state that it is an important question of language testing research whether or not tests and test items engage the learner's already familiar discourse domains and Interlanguage structures associated with them, discourse domains, for these authors, being the main types of internally created contexts. In other words, it is a question of the validity of the domain or domains chosen for the test. However, with the definition of domain provided by these authors it is difficult to have a clear picture of an "already existing discourse domain". Our best guess is then that learners will internally create *similar discourse domains* in the testing situation to the ones they had created in the *past if the conditions of the testing situation promote those similarities*. Thus, we expect that learners will perform better with familiar discourse domains rather than with non-familiar ones. In addition, Douglas and Selinker remark that "the valid test must distinguish between the Interlanguage which results from those occasions when the testee is able to nominate and control the topic, and those occasions when the testee is not." (1985: 219). To give substance to the notion of discourse domain, some examples of possible domains within the elementary-school classroom are offered next. According to FL classroom practices, in elementary school children are usually familiar with the domain of responding physically to any given command, with the domain of naming objects with the help of visuals, and with the domain of responding to questions relative to personal background. Children are expected to be familiar with the domain of posing questions to request information, with the domain of responding to questions relative to any given picture, and with the domain of repeating any given utterance. Our expectation is that a test will have more *construct validity* if these domains are chosen to be part of an oral test. I will come back to this topic later on the paper.

2.2. Level of cognitive development

So far, I have referred to language ability as being inseparable from a theory and method for instruction for the case of assessment in classrooms or language programs. What I have not mentioned yet is

that language ability goes hand in hand with the learner's stage of cognitive development. Research indicates that young children attending school continue expanding their knowledge of their native language for several more years (Cummins, 1981). Thus, learning a foreign language is built on the native language background as well as on general background about the world. According to Curtain and Pesola (1994), six year-old children can name and memorize the months in their native language but still lack of track of time. They may know their birthday date but are not able to comprehend what an hour entails. Thus, it is realistic to have similar expectations for these children when learning the foreign language.

The level of cognitive development has implications in other aspects of language including the extent to which children can express themselves orally and in writing; their capability of reflecting on language itself (known as metacognitive ability); the extent to which children can elaborate on more abstract topics; the complexity of the grammatical structures used by children; and the vocabulary inventory available to them at certain age (Genishi and Haas Dyson, 1987; Homel, Palij and Aaronson, 1987).

When assessing early FL learners orally, their stage of cognitive development must be taken into consideration, otherwise the test may overestimate or underestimate learners' cognitive abilities and, therefore, not be valid from that respect¹. I will expand on the links between the dimension of cognitive development and the many types of validity later in this section. The level of cognitive development has implications on the pragmatic dimension of the test too. In their pilot study of an oral test for Japanese learners, Carpenter, Fujii and Kataoka (1995) indicate that the way in which children respond when being tested orally is different from adults. For instance, adults would be more inclined to ask for clarification of the test directions, while children would remain silent. I will come back to this issue later.

How do we know whether a task is appropriate to children's stage of cognitive development? There are several ways of capturing learners' level of cognitive development. Once the type of test has been defined (see next section) and the objectives of the test are clear for

¹ Notice that I have not referred to language knowledge. It could happen that the examinee responds appropriately with regard to the language knowledge but be under or overestimated from a cognitive perspective. For example, asking an eight years old child, who has received several years of FL instruction, about colors and numbers from 1 to 10 may underestimate her abilities.

the tester, he or she may take a look at the course' program or syllabus, in the case of an achievement test, to develop tasks feasible for the child. Conducting class observations may also be helpful, because it might be the case that the syllabus and actual teaching differ. Observing class routines, method of instruction, teacher's prompting children, and activities in class – everything will help the examiner to construct an adequate test with respect to the level of cognitive development. Notice that it is assumed that the method of instruction is appropriate to the level of cognitive development. Furthermore, although children are grouped according to their age at school, it would be of great help to have children's individual profiles for acknowledging any possible anomaly among children, e.g. children with hearing problems or speech disorders.

2.3. Type of test

Choosing the appropriate type of test for oral assessment has direct links to all the factors shown in Figure 1, since all these factors must be present in the test for it to be valid and reliable. When specifying the domain or domains of oral test, it was mentioned that for young FL learners who only speak the target language in the classroom it would be better to construct the test based upon instructional tasks. Hence, delimiting the domain entails making decisions on the type of test as well. Tests that assess students against instructional or learning objectives are called *achievement tests*. These tests base their evaluation mostly upon information learners should have learned during the course or program, therefore these tests are based on the course syllabus or the course textbook and are administered at the end of the course. In contrast, *proficiency tests* are not based on a particular language program. Rather they are designed to test the ability of students with different language training backgrounds. In order to assess young FL learners orally in an appropriate manner we need to consider the above mentioned. If the main goal of the test is to find out whether students are showing progress at each level, an achievement test will fit the expectations of educational agents involved better than a proficiency test. Hence, learners should be assessed against the learning objectives of the level to which they belong. Other goals of the test might be to investigate whether children who started the language program early show cumulative knowledge from previous years compare to children who started the program late. Thus, students need to be assessed against learning objectives from the next lower level. In doing so, the test might indicate the existence of cumulative knowledge among children as well as differences between early beginners and late beginners of the program. It might be the case, that test designers want to know whether children are learning the foreign language through other

sources such as siblings enrolled in upper levels of the program. In this case, the test should include items from the next upper level too. Although this would not be an achievement test in the strict sense (it involves testing children against learning objectives that they have not learned yet), from the perspective of the program as a whole the test still is an achievement test.

Next, the facets of test methods according to Bachman (1990) are enumerated. According to Bachman, test method facets have the following categories: the testing environment, the test rubric,

the nature of the input the test taker receives, the nature of the expected response to that input, and the relationship between input and response. Except for the first one, the other four are presented very briefly. The rubric of the test consists of specifying test takers procedures in taking the test. It is recommended to be as explicit as possible in detailing the organization of the test, the time allocated for it as well as instructions for test takers and criteria for correctness. The second category, namely the facets of input, consists of specifying the format in which the input is presented, including channel (aural, visual), mode (receptive), form (language, nonlanguage), vehicle (live, canned) and language of presentation (native, target). In addition, the degree of speededness in which the input is presented should be specified as well as the nature of the language such as length, propositional content, vocabulary, degree of contextualization, distribution of new information, type of information, topic, and genre. It should also be included the organizational and pragmatic characteristics of the language, which in turn depend on how language ability is defined. The fourth category deals with the characteristics of the expected response. It should include the format (channel, mode, form, language of response) as well as type of response (selected, constructed). Furthermore, nature of language must be specified as well as any possible restriction on response. The fifth category, the relationship between input and response, entails assumptions of how language is encoded and decoded.

Under testing environment, Bachman includes familiarity of the place and equipment, the personnel involved in the test, time of testing and physical conditions. It is desirable to create an appropriate atmosphere where children will feel comfortable as well as encouraged to respond with long and complex utterances maximizing the amount and quality of outcomes in the target language. As mentioned above, children behave differently from adults when being tested. For instance, they do not ask for clarification of the instructions of the tasks. Carpenter, Fujii and

Kataoka (1995) report that none of the 40 children they tested in speaking "gave any deliberate explicit signals that they had not understood something the tester said. None of the children said anything like 'I don't understand' or 'Could you repeat that?' or 'Could you speak slower?' (p. 172). According to the authors, it is not clear yet whether this attitude from the children is because they do not practice such phrases in the classroom or whether it is a characteristic of the children's cognitive stage.

These findings go hand in hand with Andersen's (1990) study about young children performing different registers. She found out that children will often try a register associated with a given role, even if they have little or no knowledge about that role. She also found avoidance of features associated with certain register when the children are aware of this register but not about its features. Schachter (1974) reports the same strategy of avoidance for second language learners and Celce-Murcia (1975) cited in McLaughlin (1978) reports children avoiding words of difficult pronunciation systematically. Andersen also found that children use a particular register confidently if they know how; and "that children prefer to play the roles they know best, but once they know several fairly well, they prefer playing those with higher status." (p. 139). In another study, Genishi and Haas Dyson (1987) show that the child is sensitive to the linguistic ability and linguistic code of her listener. The findings of this study show that children switch from one language to the other according to how they perceive their listeners' linguistic ability. These findings coincide with Carrasco, Vera and Cazden's (1981) study of peer-teaching among young bilingual children. In their study, a bilingual Chicana girl talks in *Spanish* to another bilingual boy while engaging in peer-teaching. Later the girl reports in *English* to the teacher the completion of the task. In summary, all these studies show that the task of making a good test that creates the appropriate atmosphere for children with respect to the pragmatics of testing is a delicate one. Otherwise, testers may underestimate children's oral skills. Notice how children's oral skills and their level of cognitive development are both closely interrelated with the pragmatics of the testing situation.

A final comment on concerns about pragmatics of testing: these concerns are directly related to similar worries with respect to the method of instruction for young FL learners as well as with the maximization of the amount of student-talk in class (Swain 1988, 1996a, 1996b). Those concerns have no simple answers and are still motivating research on teaching methodology.

2.4. Validity issues

In this section I will elaborate on the connections between the factors discussed above and issues on validity. In this paper, I have mentioned possible threats to the validity of a test related to its content, its method and the domain being assessed (McNamara, 2000; Messick, 1989, 1996).

Messick (1989) defines validity as "an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment." (p. 13). He also calls to our attention that the test device *per se* is not valid but the interpretations of the test scores. In other words, test scores are valid if our interpretations of them reflect adequately what they are intended to mean (Messick, 1996).

Assessing children orally in a FL entails consideration of the domain, level of cognitive development, its relationship with pragmatics issues for the testing situation, and type of test discussed in previous sections. How are these factors linked to issues on validity?

First, it is necessary to establish the language ability and then the domain or construct of the oral test. It was mentioned that it would be better for young FL learners, who speak the target language only in the classroom to be evaluated against domains that are usual to the FL classroom. In a previous section, several examples of possible domains familiar to young children in the FL classroom were presented. As well as the domain of responding physically to any given command, children are also familiar with the domain of naming objects with the help of visuals, and with the domain of responding questions relative to personal background. If the test can measure accurately those domains, the test will have *construct validity*. This is because the construct (the domain or trait being measured) is well reflected on the test. Assuming that the domain is derived from the method of FL instruction, there might be three positive consequences of construct validity. One consequence is the non-existence of *construct irrelevant variance*, while another is the lack of *construct under-representation* (Messick, 1996). That is to say that the test is developed according to *how* the FL has been taught to the child. Avoiding construct irrelevant variance means eliminating all factors not pertinent to the aspect of the ability being measured. For example, when testing commands of the classroom the child should be asked to perform with physical responses. Instead, an incorrect interpretation of testing commands would be to ask the child to *say* the commands. On the other hand, lack of construct under-representation means that

tasks will not be too easy for examinees. The third aspect of good test-construction is *content validity*. The extent to which the items on the test are derived from *what* has been taught has an impact on the test being content valid (Genesee and Upshur, 1996; Bachman and Palmer, 1996; Cohen, 1994; Alderson, Clapham and Wall, 1995). Children need to be tested based upon what they have learned.

The domain is not the only factor that has implications on test construct and test method; both the child's level of cognitive development and pragmatics of the testing situation are also related to this aspect of validity. The child's level of cognitive development is linked to test construct validity to the extent that child oral skills are inseparable from stage of cognitive development as it was discussed previously. Thus, test designers have to take account of this factor as part of the domain. Similarly to how it has been explained above, there are direct links between the child's level of cognitive development and the test method. First, it is desirable to avoid construct irrelevant variance, or in other words, we do not want to challenge the child's cognitive stage. For example, first graders may be able to formulate questions but do not know the concept of 'question' itself. Therefore in order to ask first graders to formulate questions, one may want to use a role-play activity with a puppet in which the learner needs to find out the puppet's name, his age and so on. It would not be appropriate to tell the learner: "ask a question to the puppet." The purpose of this task is to measure the child's performance in formulating questions not measuring her knowledge on *what* a question is. Second, it is important to avoid easy tasks from the cognitive perspective so that there will not be construct under-representation. For instance, learners who have been several years in the language program should be able to understand and respond to longer and more complex utterances than learners who just started the language program.

The pragmatics of the testing situation set up the conditions under which the child displays her performance during the test. Thus, issues on the pragmatics of the test are related directly to both threats of validity regarding the test method, namely construct irrelevant variance and construct under-representation. As was explained previously, children have particular reactions to certain role-plays. Some practices with role-play might inhibit children from showing off their knowledge of the language. Furthermore, the pragmatics of the testing situation is closely related to another aspect of validity called *face validity*, meaning the surface acceptability to those involved in the test development (McNamara, 2000).

5.5. Reliability issues

In this section, issues on reliability related to the time before the test is administered are commented on briefly. I will not touch other reliability matters that are related to scores analysis, rater inter-reliability and other issues related to the time after the test has been administered. The focus is on five elements of test administration: number and selection of items, trial of items, test administrators, raters, and rating scales. In addition, connections among the factors discussed in previous sections and these elements are established.

The type of items included in the test should be carefully selected to avoid the possibility of students responding successfully due to guessing. Furthermore, if the test aims to compare students' cumulative knowledge across levels, there should be some items shared by every level, whereas other items are selected to test instructional objectives of one particular level only. Thus, more items are needed. In order to establish whether an item is a good item, namely that it measures what is intended to be measured, it is important to run a trial of the test. Analysis of the results of the trial will let us know about the item difficulty. These trials should be made with a representative sample of learners so that results are sufficiently trustable.

Test administrators and raters must be aware of child particular characteristics regarding both their cognitive stage and the pragmatics of the testing situation. Thus, rater training should involve familiarization on those matters. Furthermore, raters need to be familiar with child speech characteristics, including stutter, vacillation and time of pauses between utterances (Vihman, Macken, R. Miller, Simons and J. Miller, 1985; Silverman, Noa and Russell, 1976). The purpose of raters being familiar with child characteristics will increase the reliability of the results by reducing the possibility of raters rating other aspects of language ability instead of the ones that are intended to be assessed. With respect to the rating scales used, the challenge is that they have to capture child responses accurately. Following the same rationale, rating scales should be constructed taking into consideration child characteristics. Furthermore, if the test is validated using correlations with other tests that involved oral skills too, it is important to be consistent in the use of similar rating scales for comparability purposes.

3. Summary

In this paper, I seek to explore factors involved in testing children in speaking a foreign language. I have shown how these factors are all interrelated to each other. Hence, in order to understand one of them, one must understand the others. I have put particular emphasis on the factors' implications on validity issues. On the other hand, there are other issues that I have chosen not to comment on but whose implications for test construction I acknowledge too. These issues include examinee's individual differences (Herzel, 1999), the distance between the learner's native language and the target language typologically speaking (Kellerman, 1995), the nature of the test from a political perspective, and possible effects of a test on teaching, also called washback effect (Lapkin, 1985).

A last comment on the factors' interrelationship is to call attention to its dynamic character. On-going research, both theoretical and empirical, on language ability and child cognitive development as well as its implications on the pragmatics of testing situations will have an impact on other dimensions such as validity and reliability and type of test. It is a never-ending story in which all the participants will benefit from the findings.

References

- Alderson, J.C., C. Clapham and D. Wall. 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Andersen, E. 1990. *Learning to speak with style*. London and New York: Routledge.
- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L., and S. Palmer. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Canale, M. and Swain, M. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1-47.

- Canale, M. 1983. On some dimensions of language proficiency. In J.W. Oller, Jr., (ed.), *Issues in language testing research* (pp. 333-342). Rowley, MA: Newbury House.
- Carpenter, K., N. Fujii and H. Kataoka. 1995. An oral interview procedure for assessing second language abilities in children. *Language Testing* 15/2: 157-81.
- Carroll, J.B. 1968. Contrastive Analysis and Interference Theory. In J.E. Alatis (ed.), *Report of the Nineteenth Annual Round Table Meeting on Linguistics and Language Studies* 21, 113-22. Washington D.C.: Georgetown U.P.
- Carrasco, R., A. Vera and C. Cazden. 1981. Aspects of bilingual students' communicative competence in the classroom. In Druan, R., editor, *Latino language and communicative behavior*. Norwood, NJ: Ablex.
- Cohen, A.D. 1994. *Assessing language ability in the classroom. Second edition*. Boston, MA: Heinle and Heinle.
- Cummins, J. 1981. The role of primary language development in promoting educational success for language minority students. In *Schooling and language minority students: A theoretical framework*. Los Angeles: Evaluation, Dissemination and Assessment Center, California State University, Los Angeles. 3-49.
- Curtain, H. and C.A. Pesola. 1994. *Languages and Children. Making the Match*. New York: Longman.
- Douglas, D. and L. Selinker. 1985. Principles for language tests within the discourse domains theory of interlanguage: research test construction and interpretation. *Language Testing* 2 (2), p. 205-26.
- Genesse, F. and J. Upshur. 1996. *Classroom-based evaluation in Second Language Education*. Cambridge: Cambridge University Press.
- Genishi, C. and A. Haas Dyson. 1987. *Language assessment in the early years*. Norwood, NJ: Ablex.
- Herzel, K. 1999. The Relationship between Individual Differences in Receptive and Expressive Language Ability and Social Information Processing in Children. Dissertation. State U of New York, Binghamton.

- Homel, P., M. Palij and D. Aaronson (Eds.). 1987. *Childhood bilingualism. Aspects of linguistic, cognitive, and social development*. Hillsdale, NJ: Lawrence Erlbaum.
- Hymes, D. 1972. On communicative competence. In K.B. Pride and J. Holmes (Eds.), *Sociolinguistics*, pp.269-93. Harmondsworth: Penguin.
- Kellerman, E. 1995. Crosslinguistic Influence: Transfer to Nowhere?. *Annual Review of Applied Linguistics* 15, p. 125-50.
- Lado, R. 1961. *Language testing*. New York: McGraw-Hill.
- Lapkin, S. 1985. Pedagogical implications of direct second language testing: A Canadian example in K. Hyltenstam, and M. Pienemann (Eds.) *Modelling and assessing second language acquisition*. Multilingual Matters, 18. Avon: Clevedon.
- McLaughlin, B. 1978. *Second-language acquisition in childhood*. Hillsdale, NJ: Lawrence Erlbaum.
- McNamara T.F. 2000. *Language Testing*. Oxford: Oxford University Press.
- Messick, S. 1989. Validity. In Linn R. (Ed.) *Educational Measurement*. NY: American Council on Education-Macmillan Publishing Company.
- Messick, S. 1996. 'Validity and washback in language testing' *Language Testing* 13/3: 241-56.
- Rhodes, N. and L. Branaman. 1999. *Foreign Language Instruction in the United States. A National Survey of Elementary and Secondary Schools*. Washington, DC and McHenry, IL: Center for Applied Linguistics/Delta Systems.
- Schachter, J. 1974. An error in error analysis. *Language Learning* 24, 205-14.
- Silverman, R., J. Noa and R. Russell. 1976: *Oral language tests for bilingual students*. Washington, DC: US Office of Education.
- Swain, M. 1988. Manipulating and Complementing Content Teaching to Maximize Second Language Learning *TESL Canada Journal*, 6: 68-83.

-
- Swain, M. 1996a. Integrating language and content in immersion classroom: Research Perspectives. *The Canadian Modern Language Review*, 52: 529-48.
- Swain, M. 1996b. Discovering Successful Second Language Teaching Strategies and Practices: From Program Evaluation to Classroom Experimentation *Journal of Multilingual and Multicultural Development* 17, no. 2-4.
- Thompson, L. 2001. Foreign Language Assessment: 30 Years of Evolution and Change. *ERIC Vol.23, No.2*. Washington, DC. Center for Applied Linguistics.
- Vihman, M., M. Macken, R. Miller, H. Simmons and J. Miller. 1985. From babbling to speech: a reassessment of the continuity issue. *Language* 61, 397-445.