Computerized Assessment of Proficiency (CAP)

Test Specifications

Center for Applied Second Language Studies (CASLS)
University of Oregon

Contents

Introduction

This document was prepared by the Center for Applied Second Language Studies (CASLS). CASLS is a Title VI K-16 National Foreign Language Resource Center located at the University of Oregon. CASLS core mission is promoting international literacy by supporting communities of educators and by partnering with those communities to develop a comprehensive system of proficiency-based tools for lifelong language learning and teaching.

CASLS is supported almost exclusively by grants from private foundations and the federal government. Reliance on receiving competitive grants keeps CASLS on the cutting edge of educational reform and developments in the second language field. CASLS adheres to a grass-roots philosophy based on the following principles:

- Teachers are the solution, not the problem. Support them, don't preach to them.
- All children have the ability to learn a second language and should be provided with that opportunity.
- The purpose of language learning is meaningful communication.
- Meeting the needs of teachers and students is our top priority.

The Computerized Assessment of Proficiency (CAP) is an online test of proficiency developed by CASLS. Many of the test items developed for CAP are licensed by Avant Assessment and delivered operationally under the name STAMP (**STA**ndards-based **M**easurement of **P**roficiency).

Statement of Purpose

CASLS believes that assessment is an integral part of a successful language program. We call this approach "empirical pedagogy": the notion that instruction should be informed by data on student performance. A good language program will use a range of assessments to provide feedback on both the students themselves as well as the program in general. Assessment can and should take many forms, both formal and informal. One of the best sources of information about students' language ability is the professional judgment of their classroom teacher based on observations made over the course of a semester, through a variety of means, including performance assessments, achievement tests, self-evaluations, etc. This type of longitudinal observation is relatively unaffected by transitory events, such as the student having an "off" day or feeling ill the day of the test, and thus offers a more stable picture of the student's ability on a variety of occasions. Portfolio-based assessment, in which student work (e.g., compositions, audio and video recordings, collages, tests, reflections, projects, etc.) is collected over the course of the semester, is popular for this reason.

CAP is designed to provide an overall estimate of a student's proficiency in four skills. It is a snapshot of language ability based on a relatively short number of tasks. As such, the CAP is not a substitute for the judgment of an experienced classroom teacher. CAP can be used effectively, however, to gauge general proficiency at the start of a course for placement purposes or to provide an indication of general proficiency at the end of a course for summative assessment. As a standardized assessment, it can provide a common touchstone for comparison at the school, district, or state level. A foreign language instructor knows his or her students the best, but does not necessarily know how those students compare to students in similar programs in other places. A standardized assessment like CAP can help facilitate such comparisons.

Description of the test taker

The target test takers for the CAP are adult (age 13+) language learners. They are assumed to be native speakers of English or to have a high degree of fluency in English, including English literacy. The test takers will be primarily students in programs where the target language is taught, but may also be persons seeking to enter such programs, including those who have learned the language informally.

Description of the test score user

Test takers, language teachers and program administrators are the intended score users. Test takers will use the test score to evaluate their progress towards their language learning goals. Language teachers will use the scores to help inform (in conjunction with multiple other sources of information) placement decisions and summative evaluations of students. At the class level, aggregate information can help inform curricular decisions for program administrators.

Intended consequences of test score use

The ultimate goal of the test is to increase the foreign language capacity of language learners in the US. As such, it is hoped that use of the test has a positive washback on programs in terms of more highly valuing proficiency and meaningful language use over rote memorization.

CASLS suggests that educators not use CAP (or any other single assessment) as the sole basis of making decisions affecting students. These decisions might include graduation and credit issues. Used in connection with other measures, such as course grades, teacher evaluations, and other external assessments, CAP can be a useful tool for making these decisions about individual students.

Construct for the CAP

CAP can be considered a "proficiency-oriented" test. Language proficiency is a measure of a person's ability to use a given language to convey and comprehend meaningful content in realistic situations. CAP is intended to gauge a students' *linguistic capacity* for successfully performing language use tasks. CAP assesses test taker performance on language tasks in four modalities (listening, speaking, reading, writing) as evidence for this capacity. An additional Contextualized Grammar section is available for programs desiring evidence of test takers' knowledge of the linguistic system.

In CAP, realistic communication settings provide the inspiration for reading and listening tasks. In many cases, authentic materials are used with minimal adaptation for the purposes of the test. In other cases, these materials provide the template or model for materials created specifically for the test. Items are not developed to test a particular grammar point or vocabulary item (with the exception of the Contextualized Grammar section). Rather, the tasks approximate the actions and contexts of the real world, which permits making informal inferences as to how the learner would perform in the ``real world''.

Test level

CASLS reports assessment results on the CASLS Benchmark Scale. Several points along the scale have been designated as *Benchmark Levels*. These Benchmark Levels include verbal descriptions of the proficiency profile of a typical student at that point in the scale.

The Benchmark Level descriptions derive their lineage from well-known proficiency scales, notably the Foreign Service Institute / Interagency Language Roundtable (FSI/ILR) scale and the American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines. The relationship between the scales is shown in Table 1.

Table 1

ILR	ACTFL	Benchmark Scale	Benchmark Levels
3	Superior	10 Benchmark Level	Refining
2+	Advanced-High	9	Expanding
2	Advanced-Mid	8 Benchmark Level	
	Advanced-Low	7	
1+	Intermediate-High	6	Transitioning
1	Intermediate-Mid	5 Benchmark Level	
	Intermediate-Low	4	
0+	Novice-High	3 Benchmark Level	Beginning
	Novice-Mid	2	
0	Novice-Low	1	

Note that the Scale steps don't necessarily reflect equal intervals, and the amount of time needed to move from one Benchmark to the next increases in an exponential manner as one moves up the scale. Although it is tempting to take the points on the scale as objectives for a series of language classes (e.g., Benchmark 3 by the end of Year 1, Benchmark 5 by the end of Year 2, etc.), caution should be used in applying the Benchmark levels without sufficient consideration of the actual instructional time needed to make progress.

Benchmark Levels

Level 3 (Beginning proficiency): Beginning proficiency is characterized by a reliance on a limited repertoire of learned phrases and basic vocabulary. A student at this level is able recognize the purpose of basic texts, such as menus, tickets, short notes, etc. by understanding common words and expressions. The student is able to understand a core of simple, formulaic utterances in both reading and listening. In writing and speaking, the student is able to communicate basic information through lists of words and some memorized patterns.

Level 5 (Transitioning proficiency): Transitioning proficiency is characterized by the ability to use language knowledge to understand information in everyday materials. The learner is transitioning from memorized words and phrases to original production, albeit still rather limited. In reading, students at this level should be able to understand the main ideas and explicit details in everyday materials such as short letters, menus, and advertisements. In listening, students at this level can follow short conversations and announcements on common topics and answer questions about the main idea and explicitly stated details. In speaking and writing, students are not limited to formulaic utterances, but can express factual information through the manipulation of grammatical structures.

Level 8 (Expanding proficiency): Expanding proficiency is characterized by the ability to understand and use language for straightforward informational purposes. At this level, students can understand the content of most factual, non-specialized written or spoken texts intended for a

general audience, such as newspaper articles, television programs, and the like. In writing and speaking, students have sufficient control over language to successfully express a wide range of relationships (e.g, temporal, sequential, cause and effect, etc.).

Level 10 (Refining proficiency): Refining proficiency is characterized by the ability to understand and use language that serves a rhetorical purpose and involves reading or listening "between the lines". Students at this level can follow spoken and written opinions and arguments, such as those found in newspaper editorials. The students have sufficient mastery of the language to shape their production, both written and spoken, for particular audiences and purposes and to clearly defend or justify a particular point of view.

The four Benchmark Level labels can be remembered by the mnemonic "BETTER" (**BE**ginning, **T**ransitioning, **E**xpanding, and **R**efining).

Content and structure of the CAP

The present version of CAP has of five sections:

- Interpretive Reading
- Presentational Writing
- Interpretive Listening
- Presentational Speaking
- Contextualized Grammar

Although the five sections of CAP are meant to work together to give a snapshot of the student's overall proficiency, the sections themselves are scored separately and can be delivered in a modular fashion. There is no aggregate score on CAP. This is done to give programs the maximum flexibility in using the test. Programs can choose to use all sections of CAP outright or can choose specific sections to supplement assessment practices already in place.

Interpretive Reading

Reading proficiency is defined as the ability to extract information from written texts for a particular purpose. Following vanDijk & Kintsch (1983), two types of comprehension processes are assumed: local and global. *Local* processes refer to detecting information within a clause, while *global* processes refer to comprehending information across clauses. Proficient reading requires the use of both of these processes in complementary fashion. In addition, there is an interaction between the proficiency of the reader and the difficulty of a text. At lower levels of proficiency, readers will be limited to understanding learned words or phrases. Though it would be possible to present a Beginning reader with a novel and have them indicate the words and phrases that they can recognize, this would be frustrating for them and would provide more information about what they *can't* do in the language than what they *can* do. For this reason, the texts used in reading tasks at the lower level are chosen to represent those texts that are read primarily for immediate personal needs, such as signs, time tables, and short notes. As reading proficiency increases, the learner will be able to interact with a wider variety of texts and will not be as reliant on text-external information to derive meaning from them.

Text level

An important part of developing reading items is ensuring that the texts chosen for presentation are appropriate for the level in question. This is done through a process of text leveling, based primarily on the work of Child (1987, 1998). Text leveling uses the interaction of text modes, linguistic features, and Benchmark Level descriptions to appropriately identify the level of a reading passage. Test items are designed to match the level of the texts.

Typical text length

Though in some instances differences in orthography confound what constitutes a "word" in any given language, the following chart shows the approximate length of texts at various levels.

Refining: 200 – 350 words
Expanding: 150 – 300 words
Transitioning: 20 – 150 words
Beginning: 1 – 30 words

Typical types of texts

Although the actual passages for the test are chosen based on several criteria, the following are examples of the types of texts frequently found at the various proficiency levels. This is meant to be illustrative rather than comprehensive.

Refining	newspaper editorials, movie reviews, commentaries, literary excerpts
Expanding	factual descriptions such as found in newspaper and magazine articles, personal correspondence, simple business correspondence,
Transitioning	menus, schedules, simple correspondence, simple descriptions of people and places
Beginning	signs, simple menus, advertisements, short notes, postcards, shopping lists

Items per text

The number of items per text is a function of the length of the length and level of the text. Longer and more involved texts tend to support more questions than short texts.

Refining: 2 – 5 items per passage
Expanding: 1 – 4 items per passage
Transitioning: 1 – 3 items per passage
Beginning: 1 – 2 items per passage

Item Format

All items on the reading test are multiple choice, with one correct option and three incorrect "distractor" options. Reading texts are presented in the target language; reading items and options are written in English. Each text is preceded by a short description of the context written in English (e.g., "You are reading this letter from your friend.").

For passages that have more than one associated item, items are presented individually. The test taker must answer the first question to see subsequent questions. Backtracking is not permitted by the test system.

Item Functions

Items in CAP can be broadly broken down into two types: "scan for gist" (SG) and "extract detail" (ED). SG items require the test taker to grasp the main points of the text; ED items require the student to comprehend particular sections of the text. These correspond roughly to the global and local processes described above. In general, if a text has more than one item the first will be an SG item and the others will be ED items.

Scoring

Items are automatically scored correct or incorrect by the test delivery engine.

Presentational Writing

Writing proficiency is the ability to appropriately structure and produce a written message that can be successfully comprehended by an interlocutor at a later time. At lower levels of proficiency, gaps in linguistic knowledge and appropriate written conventions lead to writing that is understandable only to someone sympathetic to emerging language. At higher levels of proficiency, the writer is able to use language to craft writing that is understandable and appropriate for a wide variety of situations.

Format

Examinees read a short prompt in English that describes a situation accompanied by appropriate supporting visual material. An example prompt might be "You are arriving by plane tomorrow. Write a letter to your host family describing yourself so that they can recognize you at the airport." Examinees type their response into a text box. When they are satisfied with their answer, they click on a SUBMIT button and are taken to the next item. Because the input is done via keyboard rather than pen and paper, a reasonable amount of keyboarding skill is required for the writing portion of the test.

Scoring

Responses are scored by trained raters according to a rubric. Rubric categories include:

Rating Category	Beginning	Transitioning	Expanding	Refining
Text Type	X	X	X	X
Comprehensibility	X	X	X	X
Language Control		X	X	X
Vocabulary Use			X	X
Syntax			X	X
Discourse			X	X
Accuracy			X	X

The rubric is "expanding" in that additional rating categories are introduced as the proficiency level increases.

Text Type refers to the form of the examinee's text and ranges from words and phrases at the low end to connected paragraphs at the upper end.

Comprehensibility refers to the ease with which the reader can understand the language that the examinee is producing. At lower levels, the interlocutor is expected to be sympathetic to the writing of non-native speakers; at higher levels, the interlocutor is not assumed to have had previous contact with non-native speaker writing.

Language Control refers to the ability of the examinee to use the structures of the language to make him or herself clearly understood. The focus is on the clarity of the message rather than the absolute grammatical accuracy. At the upper levels, the Language Control category is further differentiated into *Vocabulary Use*, *Syntax*, *Discourse*, and *Accuracy*.

Responses that are completely in English or off-topic are considered unratable.

Interpretive Listening

Listening proficiency is defined as the ability to extract meaning from aurally presented material. Like reading, this involves both local and global processes. In addition, listening proficiency requires the ability to deal with the input in *real time* as it is presented (Buck, 2001). As with reading, there is an interaction between the proficiency of the listener and the difficulty of the text. Presenting a test taker with passages of increasing difficulty is one way to gauge that test-taker's proficiency.

Text level

As with reading items, texts are leveled according to the system originally pioneered by Child (1987, 1998). Test items are designed to match the level of the passages.

Typical passage length

In order to minimize the memory burden on the test takers, listening passages are kept relatively short, even at the higher levels. Test takers are *not* allowed to take notes while listening.

Refining: 40 – 90 seconds
Expanding: 40 – 90 seconds
Transitioning: 20 - 60 seconds
Beginning: 10 – 60 seconds

Items per text

The number of items per text is a function of the length of the length and level of the text. Longer and more involved texts tend to support more questions than short texts.

Refining: 2 – 5 items per passage
Expanding: 1 – 4 items per passage
Transitioning: 1 – 3 items per passage
Beginning: 1 – 2 items per passage

Format

All of the listening items are multiple-choice. Each listening passage is preceded by a short description of the context in English (e.g., ``You are in the department store when you hear an announcement over the PA system."). Items and answer options are written in English. Photographs or illustrations appropriate to the content or setting of the passage are presented.

The passage is presented *once* per item. Thus, a listening passage with only one item will be presented one time; a listening passage with three items will be presented a total of three times. This method of presentation was chosen to minimize the effect of memory for passages with multiple items.

Item Functions

Similarly to the Reading test, items in Listening CAP can be broadly broken down into two types: "scan for gist" (SG) and "extract detail" (ED). SG items require the test taker to grasp the main points of the text; ED items require the student to comprehend particular sections of the text. In general, if a listening text has more than one associated item, the first will be an SG item and the others will be ED items.

Scoring

Items are automatically scored correct or incorrect by the test delivery engine.

Presentational Speaking

Presentational speaking is the ``creation of messages in a manner that facilitates interpretation by members of the other culture where no direct opportunity for the active negotiation of meaning between the members of the two cultures exists'' (Standards, 1996, p. 34). As speakers move from low to high proficiency, their ability to use their knowledge of the components of language (e.g., lexical, syntactic, etc.) to create meaning with the language increases. Widdowson (2001) describes this process as internalizing the ``communicative potential'' (p. 19) of language. At lower levels of proficiency, the speaker will mainly rely on learned words and phrases. At this level, the interpretation of the speaker's intentions may only be possible by interlocutors accustomed to dealing with language typical of emergent speakers. As proficiency increases, the speaker will be able to use the language more productively in a manner that is understandable to a wider range of interlocutors.

Format

Examinees read a short prompt in English that describes a situation along with any appropriate supporting visual material. For example, examinees might be presented with photos of two apartments and a task of ``You and your friend are looking for an apartment to share. You have seen these ads in the newspaper. Call your friend and leave a message on his answering machine describing the two apartments''. Examinees record their responses by clicking on a RECORD icon on the screen and speaking into a microphone connected to the computer. Examinees have the option of re-recording their response if desired. Once the examinee clicks a SEND button on the screen, the response is transmitted to the server and the next item is presented. A total of three prompts are presented to each examinee.

Scoring

Responses are scored by trained raters according to a rubric. Rubric categories include:

Rating Category	Beginning	Transitioning	Expanding	Refining
Text Type	X	X	X	X
Comprehensibility	X	X	X	X
Language Control		X	X	X
Vocabulary Use			X	X

Syntax			X	X
Discourse			X	X
Accuracy			X	X

The rubric is "expanding" in that additional rating categories are introduced as the proficiency level increases.

Text Type refers to the form of the examinee's text and ranges from words and phrases at the low end to connected paragraphs at the upper end.

Comprehensibility refers to the ease with which an interlocutor can understand the language that the examinee is producing. At lower levels, the interlocutor is expected to be sympathetic to the language of non-native speakers; at higher levels, the interlocutor is not assumed to have had previous contact with non-native speakers.

Language Control refers to the ability of the examinee to use the structures of the language to make him or herself clearly understood. The focus is on the clarity of the message rather than the absolute grammatical accuracy. At the upper levels, the Language Control category is further differentiated into *Vocabulary Use*, *Syntax*, *Discourse*, and *Accuracy*.

Responses that are completely in English or off-topic are considered unratable.

Contextualized Grammar

Following Purpura (2004), grammatical knowledge is considered to consist of grammatical form (accuracy) and grammatical meaning (meaningfulness).The test assesses to what degree the test-taker has mastered the ability to recognize grammatically appropriate and meaningful utterances and texts. This section of the test is considered "optional" and is intended to supplement the proficiency information from other sections of the test with explicit measures of the students' grammatical knowledge. The primary use of this information is for placement into programs for which there is a strong grammatical component in the syllabus.

Format

Contextualized Grammar consist of target language texts presented in a realistic context. Some words or phrases in the text have been replaced with blanks. Students must choose the best answer to complete the blanks from a set of four options.

Levels

Items are classified as "Easy", "Medium", or "Hard". Initial classification is done based on the sequence in which the structure in question is usually taught in formal settings. This is adjusted based on empirical results from pilot testing. Care is taken to ensure that the level of the passage is not greater than the difficulty of the item. Thus, "Easy" grammar items would be presented in the kinds of texts that are appropriate for Beginning level students.

Scoring

Items are automatically scored correct or incorrect by the test delivery engine.

Test delivery

Algorithm

The reading and listening sections of CAP are delivered adaptively based on a multistage testlet design. In this type of delivery, testlets with items at different levels are linked together through a series of routing points. The set of testlets is called a "panel". In the first stage, the test-takers see items of medium difficulty. Based on their performance on these items, they will be routed to a more difficult, easier, or similar level of items in the second stage. Illustration 1 graphically presents this algorithm, with arrows indicating possible test paths:
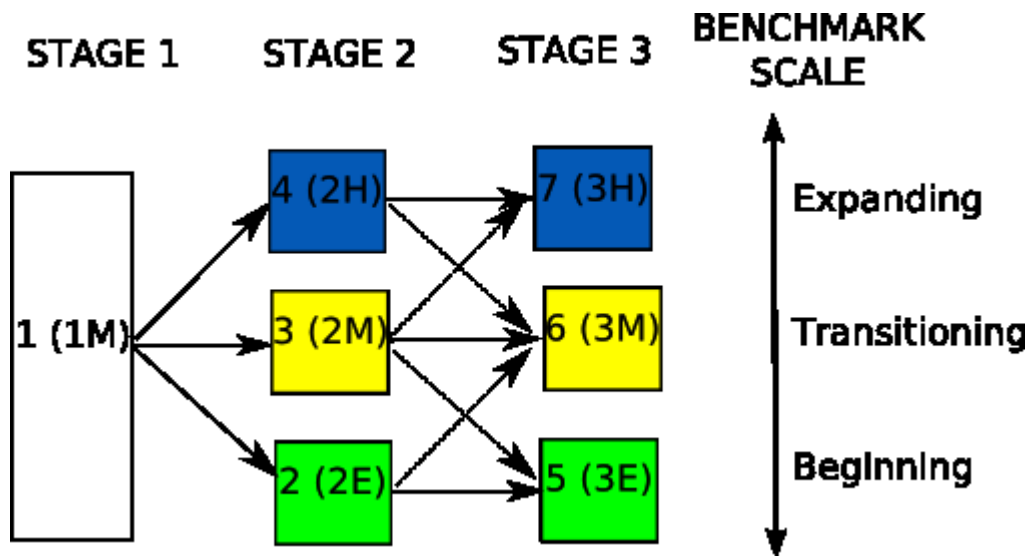
Illustration 1: Multistage Algorithm

Note that in this illustration, test items in the "Hard" testlets only cover the Expanding (ACTFL Advanced) level of the Benchmark Scale for this particular test panel. This is done to keep the test size manageable while covering the range typically found in language programs. For programs with more advanced learners, a second "test panel" is available that goes from the Transitioning to Refining range.

The Contextualized Grammar section is computer delivered, but not adaptive in the traditional sense. Because there is no need to frustrate students with items that are far too hard for them, the Contextualized Grammar section is delivered as a series of testlets with progressively more difficult items. If students do not successfully answer a minimum number of items on the testlet, the test will end. This is illustrated graphically in Illustration 2.
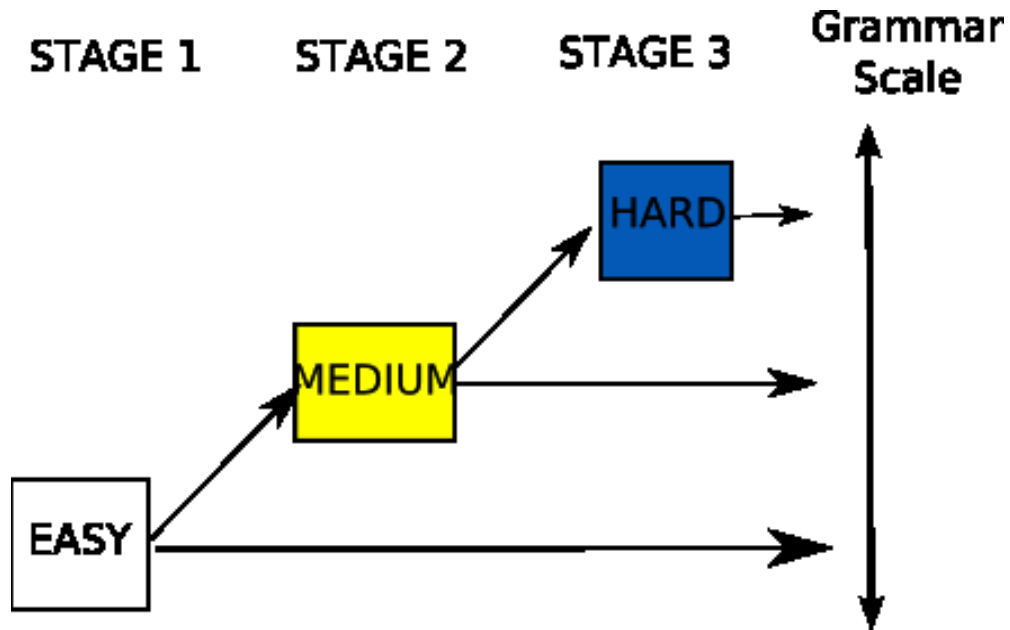
Illustration 2: Contextualized Grammar Algorithm

Test length

The overall test length is fixed, though the proportion of items that a test taker will see at any given level depends on their particular path through the stages.

The listening and reading section each have 45 items, arranged as follows:

Stage 1   15 items
Stage 2   15 items
Stage 3   15 items
Total   45 items

The Contextualized Grammar section has 45 items arranged as follows:

Stage 1   15 items
Stage 2   15 items
Stage 3   15 items
Total   45 items

The writing section and speaking section have four prompts each:

Expanding Level   1 prompt
Transitioning Level   2 prompts
Beginning   1 prompt
Total   4 prompts

Time limit

There is no explicit time limit set for this test.

Current Status

At present (November, 2008), CAP is in pilot status. Because empirical data are not yet available to create multistage panels, the test is being delivered using a "floor first" algorithm that starts at Beginning level and progresses as high as Expanding level based on the responses of the test taker. Score reporting is not yet available, but is expected in early 2009. Teachers can rate the speaking and writing samples of their own students using a modified rubric.

References

Buck, G. (2001). *Assessing listening*. Cambridge.

Child, J.R. (1987). Language proficiency levels and the typology of texts. In Byrnes, H. & Canale, M. (Eds.) *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 97-106). National Textbook Company.

Child, J.R. (1998). Language skill levels, textual modes, and the rating process. *Foreign Language Annals*, *31*, 381-391.

National Standards Project. (1996). *Standards for foreign language learning: Preparing for the twenty-first century*. Yonkers: National Standards Project.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge.

VanDijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Widdowson, H. G. (2001). Communicative language testing: The art of the possible. In Elder, C., Brown, A., Grove, E., Hill, K., Iwashita, N., Lumley, T., McNamara, T., & O'Loughlin, K. (Eds.) *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 12-21). Cambridge.