

Assessment of Advanced Language Proficiency: Why Performance-based Tasks?

Elana Shohamy & Ofra Inbar

CALPER Professional Development Document 2006

Language assessment currently offers a wide repertoire of procedures for evaluating learners' proficiency. These include the traditional multiple choice and true/false test formats, as well as alternative assessments such as portfolios, exhibitions, projects and self/peer assessments. The choice of which assessment instruments to use at any given time and given context depends on the purpose of the assessment and the content and language that the teacher wishes to assess. Similarly, assessment in advanced language proficiency courses can be carried out in varied forms, with the teacher/assessor selecting the procedures which are suited for evaluating the course objectives. In this paper we focus on one type of assessment instrument: performance-based tasks. The following sections describe what performance-based tasks are and their suitability for assessing advanced language proficiency. We then provide a number of examples of performance-based tasks specifically geared towards advanced language proficiency learners.

1. Content and Language Orientations in Language Programs

Language teaching invariably occurs within a relevant content. The emphasis on content varies however according to the nature of the program. In some language teaching programs the content will be thematically-based, that is to say that it will include a number of themes perceived as interesting and relevant to the language learners along with targeted grammar structures and skills. The content in this case serves as a context for language learning and is not viewed as the primary goal of the program. In more specific content-based programs the aim of the language course may be to acquire expertise in the target language within specific disciplines, such as law, medicine, political science and technology. These programs, like the English for Specific Purposes courses (LSP), would then be geared to serve the language needs of students studying disciplines, such as the ones we mentioned. In addition and further along the continuum content-wise are content-based settings in which the content area is fused with the language studies.

Thus a course in research methods, computer engineering or literature is taught entirely in the target language with the emphasis placed on the content. Immersion programs, where the entire learning program of study (rather than a specific course) occurs in the target language, are located at the furthest content end of this classification. Language programs can therefore be located along a continuum according to their content and/or language-oriented foci, with the possibility of allocating equal or differential emphasis to both content and language. Figure 1 below illustrates this notion.

Since teaching and assessment are viewed as integrated components in the teaching/learning process, the assessment policy and process will reflect the relative worth of the content or linguistic components in the program. Thus, in the programs where the content objective is perceived to be more significant than the linguistic features, assessment will primarily be concerned with the content, the linguistic variables deemed secondary. Conversely, if the intention is to evaluate writing skills in an academic writing course, then the content will probably be rated as less important, while the writing skills are prioritized. There exists yet a third possibility, where assessment of both language and content merit equal value and be assessed as such.

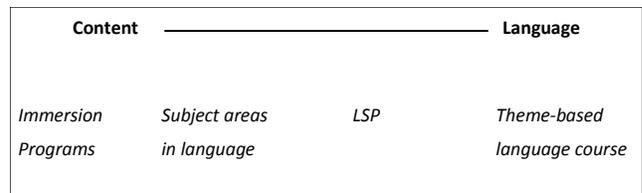


Fig. 1: The content/language continuum in language teaching programs

Although content is essentially a part of language teaching, on all levels, advanced learner classes, advanced learner classes tend to be more content focused. Students

studying in such programs are assumed to have mastered a certain level of language proficiency which enables the to cope with more advanced program objectives in terms of content and task complexity (where the learners are required to follow a number of steps in order to perform a task such as projects, thesis work and article presentation). Thus assessment of advanced language proficiency requires the use of assessment procedures which reflect the content dimension of the program as well as the acquired language skills. One such possible procedure is the *performance task*.

2. Performance Tasks

Performance tasks vary greatly and can include formats such as conducting research, presenting information orally or in written forms for various purposes, conducting a survey, simulating a radio broadcast. A more limited version of a performance task may be an open-ended item where a number of answers are deemed acceptable.

Performance tasks allow for an integration of the content and language areas acquired, thus corresponding to and serving the dual purpose of evaluating content and language (Valdez Pierce, 2002). Unlike tests which are uniform and limited in time and scope, performance tasks allow for choice and heterogeneity, measure multiple skills and are conducted over longer time periods in and out of classroom settings. Successful implementation requires the integration of different language abilities and language functions thus allowing for a broad interpretation of the learners' abilities in different areas.

Weigle and Jensen (1997) argue for performance tasks in content-based instruction (CIB). They base their discussion on the framework of test usefulness (Bachman and Palmer, 1996, p. 18). Test usefulness includes six qualities: reliability, construct validity, impact, practicality, authenticity and interactivensness. The two last qualities, authenticity and interactivensness, are seen as most relevant for content-based instruction (CBI), since "the goal of CBI is to foster language use through purposeful engagement with content [...] Assessments should be authentic in that they simulate as closely as possible the actual language use situations that students will engage in outside of the language classroom" (Weigle & Jensen, 1997, p. 207). These needs are compatible with the features of performance tasks as mentioned above. An illustration and discussion of the integration between curriculum and assessment considerations in an advanced language university program is presented in Byrnes (2002), with details available also on the Georgetown University German Department site.

Performance tasks require learners to engage in problem solving activities that require utilizing high level cognitive strategies such as analyzing, comparing and generaliz-

ing. In addition, since tasks do not presume a single correct answer they generate a variety of outcomes. Performance tasks are related to "real world" experiences and therefore necessitate the activation of prior knowledge, and encourage the authentic use of the target language as a means for accessing information in various subject areas. Performance tasks also provide opportunities for peer interaction, such as pair and group work, as well as provide learners with opportunities for reflection, self-evaluation and peer assessment.

In order to carry out a performance task such as a project, the student is required to complete a number of phases, both process and product-related, very often working as a member of a team. These include first the process phases of identifying the main issues to be dealt with, planning the structure and format of the presentation, locating and accessing information from different sources and then summarizing and synthesizing the main points. All of these process skills should be developed and utilized in the course of the program to enable students to create the final product in a project. Throughout the different stages students are encouraged to reflect on the process and on the difficulties encountered, the decisions made and lessons learnt (Huot, 1996). For a detailed discussion of performance tasks and different considerations and facets of performance assessment see Bachman, 2002; Byrnes, 2002; Norris & Yoshikawa, 1998; McNamara, 1996.

3. Designing Performance Tasks

Designing performance tasks follows a similar process used for designing other assessment procedures as described in the CALPER Professional Development Document on "The Assessment Process: A 'Multiplism' Perspective" and these include setting the purpose of the assessment, specifying the skill(s) to be assessed, designing signing the assessment instrument and scoring criteria, administering signing the assessment instrument and scoring criteria, administering the assessment tool, determining the quality of the language produced, assessing the quality of the procedures and interpreting and reporting the results (Shohamy & Inbar, 2006; Shohamy, 1998). However, since the task aims to assess both the content and language aspects, the assessment process needs to relate to BOTH components at all phases of planning, constructing and scoring as well as for the examination of the quality of the assessment, etc.

The first step in designing the task requires clarification and specification of the purpose for assessment. Specifically, what decisions need to be made based on the assessment results such as, acceptance to language programs, diagnosis of the language in relation to certain jobs and

grading at the end of a course. The second stage requires specification of the knowledge and skills in both the content and language. These then need to be presented in a scoring rubric which is described in section 4 below.

Performance tasks are often complex and multi-phased. This means that carrying out a performance task requires the learner to go through a number of steps in order to accomplish the final goal. In doing so students will need to utilize particular process skills which they might not be familiar with. If, for example, the task necessitates an internet search using search engines, some students may first have to go through the process of acquiring computer search skills. The same would be true for conducting a survey, synthesizing and presenting information, interviewing, and reporting, or any other skills that they may not be familiar with even in their first language. The task will thus include careful guidelines and templates for each phase allowing the students to acquire or brush up on the particular process skill required. An example of focusing on a process skill as part of a performance task can be found in the guidelines for a project about conflict (Baumgarten, 2003). In this multi-phased project students are asked to examine a conflict, summarize the information and then express their opinion about that conflict in a letter to a newspaper editor. The instructions for writing a summary are made available to the students as part of the guidelines for the project (<http://www.baumgartenet.com/conflicts.html>). In addition, it is vital that the students learn to reflect on the process and draw conclusions as to the various phases and to their ability to accomplish them. The reflection will relate to particular difficulties encountered in both the content and language domains and to what was learnt as a result.

As the process is long and multi-phased clear instructions are needed to allow for planning and monitoring. These can be in the form of a checklist which accompanies the task and specifies each phase and allows both the teacher and students to keep track of where the students are at a particular point in time. The scoring rubric should be made available prior to conducting the task to ensure that the objectives are transparent to the students in terms of goals and process, and that the learners know what needs to be done and expected in terms of quality in every stage along the way.

4. Scoring Performance-based Tasks—Rubrics

Since performance tasks are designed to measure both language ability and content knowledge, scoring the task is done according to the criteria which relate to both dimensions specified in a scoring rubric. A rubric is a scoring device used to indicate the learner's ability level on a specific task according to a number of pre-defined task-relevant criteria. The rubric provides the students and teachers with

a clear picture of what the student is expected to perform and on the areas that need improvements. The rubric is composed of hierarchical descriptors or indicators of performance for both content and language skills and provides the student with information regarding the specific performance criteria required for the task. It is also possible to use two different rubrics: one which relates specifically to the content (i.e. in social studies, literature, computer sciences), and another rubric which focuses specifically on the language skills being assessed. For example: in a content-based French history class students were asked to write an encyclopedia entry on the "Declaration of the Rights of Man and Citizen" by the French Assembly in 1789. The dimensions for assessing the two areas, content and language were presented in two rubrics: while in history the rubric included criteria such as familiarity with facts and historical processes, the language variables related to the genre (an encyclopedia entry) and to using appropriate style, vocabulary and language structures.

Some of the indicators in a rubric may be common to a number of tasks. In a presentation task, where a student is expected to present a topic to an audience, some generic language dimensions such as fluency, accuracy of speech and vocabulary knowledge will probably be included. However, one of the unique features of a scoring rubric is that it is designed and geared specifically for the task at hand. Thus in the case of an oral presentation, specific dimensions which relate to the subject area and the intended performance (for example explaining particular terms related to the content clearly) will be included among the indicators as well as the traits identified as typifying oral presentations in general.

5. Examples of Integrated Content and Language Performance Tasks

A collection of a large number of performance-based tasks is presented in Birenbaum (1999), including guidance from teachers as to planning, delivering and analyzing tasks. The specific process, on-going deliberation, as well as the rubrics used for scoring, offer valuable insight into specific examples of the content-based assessment process. Some of the tasks are content-oriented and require combined expertise in more than one subject area such as in history, social science, math, etc.. Two of the examples which will be described below combine a shared focus of content and language studies integrating the content studied in the history class with the language elements acquired in the English as a Foreign Language (EFL) course.

Example 1: Immigrating to the “New World” and Letter Writing

(Daskolevitch & Chen, 1999): In their English lesson students studied how to differentiate among and select suitable genres for various purposes, such as formal or informal letters. In the social studies class the students learnt how to draw on historical resources (documents, maps, diaries) to analyze historical facts. The specific performance task was to write a letter from someone living in the “New World” in the 16th century to potential immigrants living in the “Old World”, but wishing to immigrate to the “New World”. Thus the objectives were as follows:

(a) In Social Studies (history); synthesizing information from different sources; understanding the concept of “The New World” by differentiating between risks and opportunities; developing analytical historical thought, i.e., reaching conclusions on the basis of facts and expressing well-grounded coherent personal opinion.

(b) In English: Correct genre use of an informal letter; appropriate lexicon for discussing discovery expeditions and presenting convincing arguments; using correct grammatical structures and writing conventions.

The Process: In groups students received four historical sources about the ‘New World’ (2 diary entries, a picture, two maps) and were asked, as if they were residents of the US in the 16th century, to write a letter in English to a friend living in Europe. In the letter they needed to describe the situation in the New World and provide clear and well-grounded arguments as to whether they would recommend to their friend to immigrate to the ‘New World’ based on the information at their disposal.

The scoring rubric for such a task may include criteria for scoring the objectives in both the content (social studies) and language (EFL) presented in three hierarchical levels. The relevant dimensions in history will be (i) effective use of historical sources, (ii) demonstrating a historical per-

HISTORY			
Criteria	Level 1	Level 2	Level 3
Effective use of historical sources (documents, maps, pictures, journals)	Makes effective use of all the different historical sources locating, analyzing, & evaluating the evidence presented	Uses only some of the historical documents effectively	Does not use the historical sources effectively
Demonstrates historical perspective	Perceives events from the perspective of the historical period considering pertinent circumstances such as available resources, needs, and the political situation in relevant areas	Demonstrates partial ability to perceive the events from a historical perspective with regard to the pertinent circumstances	Demonstrates limited ability or no ability to perceive the events from a historical perspective
Synthesizes the information and identifies key arguments	Synthesizes the information from all the different sources to identify key arguments. The arguments presented are coherent, logically and sequentially organized	The synthesis of information from the different sources is lacking: it does not include all of the sources; only some key arguments are identified; the argument presentation lacks coherence and/or logical and sequential organization	Does not manage to synthesize the information from the different sources; key arguments are not identified; the argument presentation is incoherent.

ENGLISH AS A FOREIGN LANGUAGE			
Criteria	Level 1	Level 2	Level 3
Genre suitability	Follows the conventions of the genre (letter writing) in terms of format, register	Partially follows the conventions of the genre in format and/or register	Does not follow the genre conventions in both format and register
Using appropriate vocabulary	Uses vocabulary appropriate to the task, e.g. for providing convincing arguments, and for describing the local condition and the expedition	The vocabulary used is partially appropriate to carry out the required task	Lacks appropriate vocabulary to carry out the required task
Accuracy (sentence structure, grammar, spelling, punctuation)	The language sample is generally accurately written. Occasional errors do not impede comprehension	The language sample include some systematic errors. Comprehension is possible in most cases	The language sample includes many systematic errors which often obscure comprehension

spective and (iii) synthesizing the information and identifying key arguments. The language dimensions criteria would include (i) suitability to the genre, (ii) using appropriate vocabulary and (iii) accuracy. All the criteria are presented in the rubric above (adapted from Birenbaum, p. 143-144).

Example 2: Creating a Newspaper on the Industrial Revolution

Students were asked to compose various newspaper genres (an editorial, articles, letters to the editor, etc.) based on the factual data for a special 19th century newspaper issue on the Industrial Revolution (Chatzuel, 1999). The purpose of assessment was to determine if students could recognize and produce a written sample that was appropriate to the journalistic genres utilizing as content the topic studied in history class.

The assessment criteria provided related to English and included ‘accuracy’, ‘fluency’ and the ‘personal voice’. Had the other set of criteria in history been provided it would have probably related to factual knowledge about the period, and comprehension of historical processes. The scoring of the task was carried out by three raters: the English teacher and two of the English teacher colleagues, in order to check whether the rubrics yield similar results in terms of inter-rater reliability (i.e., agreement among the different raters). An additional rater could have been the history teacher. Students were also asked to provide feedback on their experiences in completing the task.

6. Guidelines for Designing a Performance Task and a Sample Task

The following are guidelines for planning content-based tasks as well as a sample task designed according to the guidelines:

- **Purpose:** State the purpose for assessment (both content and language): What will it be used for (diagnostic purposes, following achievements, placement, etc.).
- **Aspects to be assessed:** Decide what particular aspects will be assessed in both the language and the content areas, and whether both elements are equally important.
- **Process and product:** Relate to both the process and product facets in the task.
- **Expected products:** What do you expect the learners to come up with: a presentation, a poster, an essay, a dialogue, etc.
- **Prerequisites:** Determine the content knowledge and the language skills that the students need in order to carry out the task. Do certain knowledge and/or skill components need to be reviewed? How will they be integrated into the task?

- **Individual/ group/ pair:** How will the task be conducted? (If pair or group: individual division of labor needs to be specified).
- **Student choice:** Will the students be able to choose the topic? The manner of presentation? Whether to work individually or in a group?
- **Resources:** What resources will be available to the student: internet facilities, reference books, dictionaries, experts, assistance (by the teacher, computer experts, peers).
- **Place:** Where will the task be done – at home, in class, in the library or computer lab?
- **Time:** How much time will be allotted to completing the task?
- **Rating procedure:** If a rubric is used - what criteria will be included? How many levels will it comprise?
- **Assessors:** Who will take part in the assessment process: teacher(s), the student, other students, other raters.
- **Reflection:** Reflect on the process (planning, administration and analysis). Include tools (such as questionnaires) to collect data on student reflections.

Sample Performance Task:

Subject areas: Foreign languages and a Research Methods class

The Task: Conducting a survey on the perceptions of various language speakers, both in and outside the US, on the study of their native language in the US. Students will conduct a survey among speakers of the language they are studying (French, German, Arabic, Japanese, Korean, Russian, Spanish, etc) asking the language speakers how they perceive the teaching of their respective languages in the US (importance, success rate, attitudes of students in the US towards the language, culture issues, suggestions for topics and ideas for teaching methods etc.). The research will be conducted on the internet, using questionnaires designed and created by the students. The research process will utilize skills acquired in the research methods class. The research will be carried out and presented in the students’ target language.

Purpose: To assess ability to conduct and present a survey in the target language

Aspects to be examined: In the research methods class: conducting a survey, choosing a sample, designing the data collection instrument, analysis and presentation of the questionnaire data using suitable data analysis programs. In the target language: formulating questions, reporting the findings using language and discursal features suitable for

the report genre.

Expected products: A document in the target language describing the survey results in both verbal and graphic formats; an appendix with the questionnaire used to elicit the information. Prerequisites: determine the research methodology and the language skills required for performing the task.

Resources available to the student: Dictionaries, email and internet facilities, tools, computer software such as Excel programs, internet facilities.

Place: Class and computer lab.

Individual/ group/ pair: Individual or pair work (students' choice).

Time: Two weeks (2 lessons each week).

Rating procedure: Rubric which includes the criteria specified or each subject area. (conducting a survey, choosing a sample, designing the data collection instrument, analysis and presentation of the questionnaire data using suitable data analysis programs; formulating questions, reporting the findings using language and discursal features suitable for the report genre)

Assessors: Research methods and language instructors; self-assessment by the students and peer assessment.

Reflection: feedback questionnaires for students and instructors

In this document we argued for the suitability of performance tasks for assessing advanced language proficiency as this procedure allows for the integrated assessment of content and language in content-based programs characteristic of advanced language classes. We described the features of performance tasks and the rubrics used to assess them and illustrated this via a number of examples. We would like to point out, however, that performance tasks should not be the only means for assessing advanced proficiency and other forms of assessment can be used for this purpose as well. The choice of the assessment procedure will depend on the objectives of the assessment, the assessed content and the assessment context.

References

- Aschbacher, H., & Winters, L. (1992).** *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum.
- Bachman, L. F. & Palmer, A. S. (1996).** *Language testing in practice*. (Chapter 2: Test usefulness: Qualities of language tests). Oxford: Oxford University Press.
- Bachman, L. (2002).** Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453-476.
- Baumgarten, O. [n.d].** <http://www.baumgartenet.com>
- Birenbaum, M. (1997).** *Alternatives in assessment*. Tel Aviv: Ramot (in Hebrew).
- Byrnes, H. (2002).** The role of task and task-based assessment in a content-oriented collegiate FL curriculum. *Language Testing*, 19 (4), 419-437.
- Chatzuel, R. (1999).** An interdisciplinary task in History and English: an article from the Industrial Revolution. In Birenbaum, M. (ed), *Mindful assessment: from theory to practice*. Tel Aviv: Ramot (in Hebrew)
- Daskelovitch, L, & Chen, E. (1999).** An interdisciplinary task in History and English: A letter from the US to a friend in Europe wishing to emigrate to this country in the beginning of the 16th century. In Birenbaum, M. (ed) *Mindful assessment: from theory to practice*. Tel Aviv: Ramot (in Hebrew).
- Georgetown University, Department of German,** Developing Multiple Literacies" Curriculum Project , <http://www1.georgetown.edu/departments/german/programs/undergraduate/curriculum/writing/37047.html>
- McNamara, T. F. (1996).** *Measuring second language performance*. London: Longman
- Norris, J., Brown, J.D.,Hudson, T., & Yoshioka, J. (1998).** *Designing second language performance assessments*. (Technical Report #18). Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Center.
- Shohamy, E. (1995).** Performance assessment in language testing. *Annual Review of Applied Linguistics*, 15, 188-211.
- Shohamy, E. (1998).** Evaluation of learning outcomes in second language acquisition: A multiplicity perspective. In Byrnes, H. (ed.), *Learning foreign and second languages: Perspectives in research and scholarship*. New York: The Modern Language Association.

Shohamy, E., & Inbar, O. (2006). The language assessment process: A “multiplism” perspective. (CALPER Professional Development Document 0603). University Park, PA: The Pennsylvania State University, Center for Advanced Language Proficiency Education and Research. **Available online at:** <http://calper.la.psu.edu/publication.php?page=pdd5>

Turner, J. (1997). Creating content-based language tests: Guidelines for teachers. In Snow, M.A. & D.M. Brinton, (eds.), *The content-based classroom: Perspectives on integrating language and content*. New York: Longman.

Weigle, S.C., & Jensen, L. (1997). Assessment issues for content-based instruction. In In Snow, M.A. & D.M. Brinton, (eds.), *The content-based classroom: Perspectives on integrating language and content (pp. 201-212)*. New York: Longman.

Valdez Pierce, Lorraine (2002). Performance-Based Assessment: Promoting Achievement for English Language Learners. ERIC/CLL News Bulletin, Volume 26, No. 1. **Available online at:** <http://www.cal.org/ericcl/News/2002fall/performance.html>

More about developing performance-based tasks and rubrics:

Moskal, Barbara M. (2003). Recommendations for developing performance assessments and scoring rubrics. *Practical Assessment, Research, and Evaluation, 8 (14)*. **Available online at:** <http://pareonline.net/getvn.asp?v=8&n=14>.

More about rubrics at the following sites:

CALPER Web Site:

<http://calper.la.psu.edu/assessment/project.php>
<http://calper.la.psu.edu/projectwork/evaluation.php>

Evaluating WebQuests:

<http://webquest.sdsu.edu/webquestrubric.html>

PALS (Performance Assessment for Language Students), Fairfax County Public Schools:

<http://www.fcps.edu/DIS/OHSICS/forlang/PALS/rubrics>

Teacher Planet—Rubrics for Teachers:

<http://www.rubrics4teachers.com>

Debbie Rollins Web Site—TechTrekks:

<http://www.techtrekkers.com/rubrics.html>

Kathy Schrock’s Guide for Educators:

<http://schools/discoveryeducation.com/schrockguide/assess.html>

CAPS (Consortium for Assessing Performance Standards) Rubric, Foreign language educators of New Jersey:

<http://flenj.org/CAPS/?page=147>

SCORE (Schools of California Online Resources):

<http://www.sdcoe.k12.ca.us/score/actbank/trubrics.htm>

This CALPER Professional Development Document was produced with funds from a grant awarded to CALPER by the U.S. Department of Education (CFDA 84.229, P229A020010). However, the contents do not necessarily represent the policy of the Department of Education, and one should not assume endorsement by the Federal Government.

Revised October 2010:

URLs were updated

Please cite as:

Shohamy, E., & Inbar, O. (2006). Assessment of advanced language proficiency: Why performance-based tasks? (CALPER Professional Development Document 0605). University Park, PA: The Pennsylvania State University, CALPER.

© 2006 CALPER. All rights reserved.

Center for Advanced Language Proficiency Education and Research
The Pennsylvania State University
5 Sparks Building
University Park, PA 16802
<http://calper.la.psu.edu>