



*Center for Applied Second
Language Studies (CASLS)*

**Urdu Reading and Writing Assessment Project
Final Report**

Project Period:

8/15/2006 – 12/31/2008

Submitted on:

January 26, 2009

Prepared by:

Sachiko Kamioka

Center for Applied Second Language Studies

5290 University of Oregon

Eugene, OR 97403

Prepared for:

South Asian Language Resource Center (SARLC)

Table of Contents

INTRODUCTION	3
PROJECT PERSONNEL	3
DESCRIPTION OF THE ASSESSMENT.....	3
<i>Content and structure of CAP</i>	4
<i>Description of the test taker</i>	4
<i>Description of the test score user</i>	4
<i>Intended consequences of test score use</i>	5
CONSTRUCT OF CAP.....	5
TEST LEVEL.....	5
<i>Benchmark levels</i>	6
<i>Language ability assessed by CAP (based on Bachman & Palmer, 1996)</i>	7
TEST DEVELOPMENT	9
GENERAL OVERVIEW	9
MILESTONES	9
TASK SPECIFICATION DEVELOPMENT	10
ITEM DEVELOPMENT	10
GRAPHICS DEVELOPMENT.....	10
INTERNAL REVIEW AND REVISIONS.....	10
EXTERNAL REVIEW.....	10
ITEM COUNT	11
PILOT TESTING	12
FUTURE DEVELOPMENT	13

INTRODUCTION

With support from the South Asian Language Resource Center (SALRC) at the University of Chicago, the Center for Applied Second Language Studies (CASLS) at the University of Oregon has developed Urdu Computerized Assessment of Proficiency (CAP) for reading and writing skills. CAP is a Web-based assessment built to benchmarks characterizing proficiency levels that are compatible with ACTFL and ILR scales. Results show student proficiency from Novice-Low (ILR 0) through Intermediate-High (ILR 1+).

Urdu CAP is based on the successful model CASLS developed and implemented to create CAP in thirteen other languages. Many of the test items developed for CAP are licensed by Avant Assessment and delivered operationally under the name STAMP (Standards-based Measurement of Proficiency). A number of programs across the country have already adopted STAMP, and 50,000 students took the assessment during the 2007-2008 academic year. Urdu STAMP will be the first national standards-based proficiency assessment of Urdu. CASLS designed this reliable, accessible, and affordable test to measure student proficiency, to clarify instructional goals and outcomes, and to systematically improve Urdu language education.

Project Personnel

The following personnel have participated in the project.

SALRC:	Steven Poulos, SALRC director
Principal Investigator:	Carl Falsgraf, CASLS director
Project Coordinator:	Sachiko Kamioka, CASLS assistant director
Assessment Writers:	Qamar Jalil, lecturer, University of Wisconsin, Madison Tahira Naqvi, lecturer, New York University
Assessment Reviewers:	Fauzia Farooqui, lecturer, University of Michigan Naseem Hines, preceptor, Harvard University Shahnaz Hassan, lecturer, University of Texas
CASLS Assessment Specialists:	Martyn Clark, assessment director Krystal Sundstrom-Hebert, test developer
CASLS Pilot Coordinators:	Zahra Foroughifar, graduate research fellow
CASLS Technology Experts:	Laurence Gellert, educational software architect Charles Hollands, IT director
CASLS Urdu Assistant:	Mohammad Ali Abbasi, student worker
CASLS Graphic Assistants:	Sherry Xue Li, student worker Michelle Drenker, student worker Ivan Mirolyubenko, student worker

Description of the assessment

Urdu CAP is delivered through the Internet without the need for any special software. It

provides a general overall estimate of a student's proficiency in interpretive reading and presentational writing skills and presents a snapshot of literacy skills based on a relatively short number of tasks. As such, CAP is not a substitute for the judgment of an experienced classroom teacher. CAP can be used effectively, however, to gauge general literacy proficiency at the start of a course for placement purposes or to provide an indication of general literacy proficiency at the end of a course for summative assessment. Because it is consistent with the widely used ACTFL and ILR proficiency scales, 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.

Content and structure of CAP

Urdu CAP consists of two sections:

- Interpretive reading
- Presentational writing

The reading section consists of multiple-choice items, and the test engine scores items automatically. In the writing section, the computer captures student performance data and saves it to a database for later human scoring. Sections are scored separately and can be delivered in a modular fashion. There is no aggregate score on CAP. Programs can choose to use both sections of CAP or one section to supplement assessment practices already in place.

Description of the test taker

Adult (age 13+) language learners serve as the target audience for this test. The test takers are assumed to be native speakers of English or to have a high degree of fluency in English. Literacy in English is assumed. The test takers will be primarily students in programs where Urdu 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

Examinees, language instructors, and program administrators are the intended score users. Examinees will use the test score to evaluate their progress towards their language learning goals in Urdu literacy skills. Language instructors will use the scores to help inform (in conjunction with other sources of information) placement decisions and summative evaluations of students. Program administrators can use the aggregate information at the class level to inform curricular decisions.

Intended consequences of test score use

The ultimate goal of the test is to increase the foreign language capacity of language learners in the U.S. As such, CASLS hopes that use of the test has a positive washback on programs, leading to a higher value on proficiency and meaningful language use over rote memorization.

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

Construct for 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 gauges a students’ linguistic capacity for successfully performing language use tasks. In the reading section, genuine materials provide the inspiration for test-taker tasks. In many cases, CASLS adapted authentic materials for test purposes. 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. Rather, the tasks approximate the actions and contexts of the real world to make 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. CASLS derived the benchmark level descriptions from well-known proficiency scales, notably the FSI/ILR scale and the ACTFL Proficiency Guidelines. The relationship between the scales is shown in Table 1.

Table 1. Comparison of proficiency scale levels

ILR	ACTFL	CASLS Scale	Label
1+	Intermediate-High	6 Benchmark Level	Transitioning
1	Intermediate-Mid	5 Benchmark Level	
	Intermediate-Low	4 Benchmark Level	
0+	Novice-High	3 Benchmark Level	Beginning
	Novice-Mid	2 Benchmark Level	
0	Novice-Low	1 Benchmark Level	

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 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), users should apply caution when applying the benchmark levels without sufficient consideration of the actual instructional time needed to make progress.

CASLS' scale extends to 10 benchmark level (ILR3, ACTFL Superior). Urdu reading and writing CAP measures student proficiency between Novice-Low and Intermediate-High.

Benchmark Levels

Level 3 (beginning proficiency): A reliance on a limited repertoire of learned phrases and basic vocabulary characterizes beginning proficiency. A student at this level recognizes the purpose of basic texts, such as menus, tickets, and short notes, by understanding common words and expressions. The student understands a core of simple, formulaic utterances in reading. The student communicates basic information through lists of words and some memorized patterns.

Level 5 (transitioning proficiency): The ability to use language knowledge to understand information in everyday materials characterizes transitioning proficiency. 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 writing, students are not limited to formulaic utterances, but can express factual information through the manipulation of grammatical structures.

Language ability assessed by CAP (based on Bachman & Palmer, 1996)

		Beginning	Transitioning	Expanding	Refining
Grammar	Vocabulary	Knowledge of limited number of common words and cognates Graphology	knowledge of some general purpose vocabulary	knowledge of most general purpose vocabulary and common cultural references	knowledge of general purpose and some specialized vocabulary
	Syntax	little productive ability, but may be able to recognize memorized chunks	Familiarity with basic syntactic structures, but not complete accuracy; may be confused by complex structures	Facility with most basic syntactic structures and common complex constructions	Generally able to understand all but the most complex syntactic structures
Text	Cohesion	little or no cohesion	Some knowledge of cohesion, but may be confused by basic relationships	Ability to recognize most common relationships (temporal, sequential, cause and effect, etc.)	able to understand wide range of cohesive devices
	Rhetorical Organization	loose or no structure	loose or clear structure	able to recognize clear, underlying structure	Ability to recognize structure of argument
Pragmatic	Functional?	ability to recognize basic manipulative functions (greetings, commands, etc.)	ability to understand basic manipulative and ideational (descriptions) functions	heuristic (language for learning)	imaginative (language used to create imaginary world, poetry)

		Beginning	Transitioning	Expanding	Refining
	Sociolinguistic	combination of natural and contrived language	combination of natural and contrived language	mainly natural language	Ability to recognize register differences Figures of speech
Topical Knowledge	Not assessed, though students are assumed to have general knowledge of the world				
Strategic Knowledge	Not assessed, though some students may be able to make use of such knowledge, such as test-taking skills				

²Functional Knowledge – knowledge needed to interpret the intentions of the language user

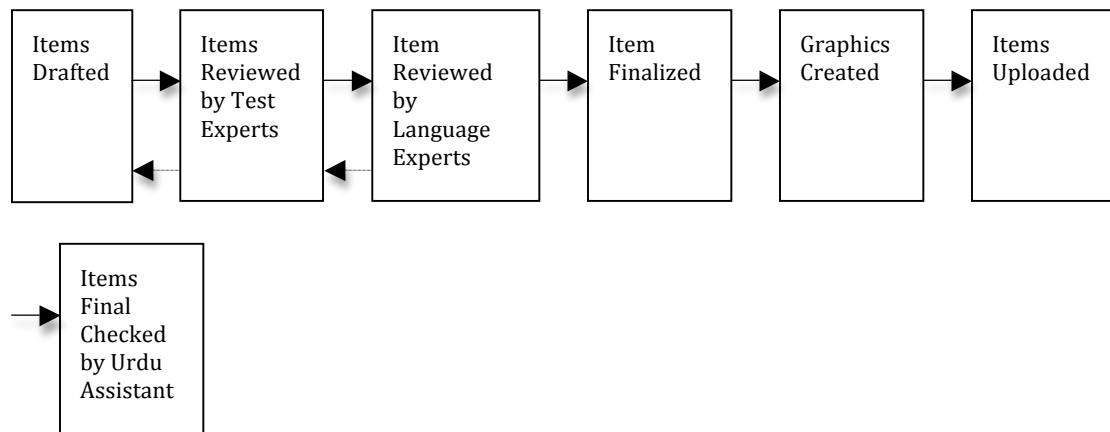
- Manipulative – using the language to affect the world around us
- Ideational – using language to express or exchange information about ideas, knowledge, or feelings
- Heuristic – using language to extend our knowledge of the world around us (for teaching and/or learning)
- Imaginative – using language to create an imaginary world (jokes, figurative language, etc.)

As the Benchmark levels increase, test-takers are expected to show evidence of increasing language ability, as indicated in the table. Reading and listening tasks in CAP are a combination of a level-appropriate texts and ability-appropriate items.

Test Development

General overview

Although any test development process involves some amount of recursion, the general item development process for reading items is illustrated in the flowchart below.



Each Urdu item went through the extensive review process illustrated above. Items were written in batches. After the writers drafted items for a batch, the coordinator reviewed the items and provided feedback on a conference call. After the writers revised items based on the discussion, they then sent items to the language experts (reviewers). The language experts reviewed each item based on a set of criteria. Upon receipt of their feedback, the writers further revised items and sent them back to CASLS. The coordinator then finalizes items with help from CASLS test developer. CASLS student workers created graphics and uploaded assessment items onto the pilot site. Finally, the writers and Urdu assistant conducted the final check.

Milestones

The table below indicates Urdu CAP test development milestones.

Time	Milestones
August 2006	Project began
October 2006	Item writing training in Eugene, OR
November 2006	First batch drafted; conference call review session
December 2006	Second batch drafted; conference call review session
June 2007	Language experts (reviewers) finished reviewing first batch; third batch drafted; conference call review session
August 2007	Conference call review session
October 2007	Item re-leveling and Novice-Low item development completed by CASLS; conference call review session
March 2008	Revisions completed and formatting issues solved

May 2008	Reviewers finished reviewing the rest of items
June 2008	Items finalized; graphics development and item uploading completed; final checked; test-run completed; pilot began
November, 2008	Writing items developed and uploaded
Present	Pilot continues

Task specification development

General test specifications for CAP/STAMP were developed based on the STAMP 2.0 framework documents. The complete test and task specifications are available on the CASLS Web site (<http://casls.uoregon.edu/papers.php>). CASLS updated existing item writing guides and job aids as necessary.

Item development

Item writing followed the conventions outlined in the CASLS item writers' guide and task specifications. Because of the iterative nature of test development, CASLS updated and expanded these documents throughout the project.

Graphics development

Because the test is compatible with any computer, CASLS renders Urdu text as a graphic to avoid font display issues when the test is delivered. For each text on the test, CASLS graphic artists imported the original text into context appropriate images, which they then uploaded to the test delivery system. The right-to-left nature of the Urdu text sometimes created formatting difficulties when transferring text and a number of discussions between the writers, Urdu assistant, and graphic artists were sometimes needed before CASLS finalized the text in the image.

Internal review and revisions

Throughout the item development process, CASLS subjected items to internal review. CASLS staff reviewed each item to ensure item quality, variety, and congruence with task specifications.

External review

Three Urdu specialists reviewed items and provided feedback. The purpose of the review was to have independent experts review the quality of the items and to ensure equal representation of Urdu spoken in India and in Pakistan. External reviewers used the following criteria:

Urdu text:

Is the text authentic?

Does the text fit to the life experiences of our target audience?

Question:

Are the questions based on important, not trivial, content?

Are all the distracters plausible?

Is there only one correct answer?

Are questions independent?

Level assignment:

Is the item level appropriate?

Bias check:

Does the item exclude offensive languages and jargons?

Item count

The table below shows the number of items written, retired (not selected), and finalized. CASLS developed two tasks each at the Novice and Intermediate levels for writing.

<i>Number of Assessment Items</i>	<i>Reading</i>	<i>Writing</i>
Written	234+	4
Retired	54+	0
Finalized	180	4

PILOT TESTING

CASLS began a pilot study in June 2008 to verify assessment item performance.

As of January 12, 2009, a total of 24 students (6 elementary, 9 third year, and 9 fourth year) from SASLI Institute have participated in the Urdu pilot. These students took the test between June 19, 2008, and July 31, 2008.

CASLS implemented a new recruitment strategy in October 2008. The pilot coordinator sent an invitation to Urdu programs and asked the writers, reviewers, and past pilot teachers to participate in the pilot. Most recently in January 2009, CASLS sent an invitation to the following institutions:

- Columbia University
- Cornell University
- New York University
- North Carolina State University
- The University of Texas at Austin
- University of California, Berkeley
- University of Chicago
- University of Michigan
- University of Pennsylvania
- University of Virginia
- University of Washington
- University of Wisconsin-Madison

CASLS recently received a response from North Carolina State University and New York University instructors and expects these universities to participate in the pilot.

FUTURE DEVELOPMENT

CASLS plans to make the test available in pilot form throughout the 2009 calendar year. Empirical data from this pilot will be used to finalize the assessment. CASLS hopes to work with SARLC to transfer Urdu CAP to Avant Assessment, which will ensure the assessment's sustainable distributions.