

How Should Student Learning Be Measured?

SALRC workshop on Assessment

Ursula Lentz
CARLA, University of Minnesota
lentz003@umn.edu

Assessment vs. Testing

Traditionally we:

Lesson/lecture/

in class practice

Assignment

quiz

test

EXAM

Workshop Objectives

- 1) Explore and understand general and classroom assessment principles and models
- 2) Discuss and identify standards, guidelines; align assessment with curricular goals
- 3) Apply knowledge to develop new /modify current assessment or assessment framework

Today

Assessment Concepts and Theory

- Matching assessment to purpose

Identifying and defining expectations

- Guidelines, criteria, standards
- Assessment concepts and types – test methods

Assessing for Proficiency

- Rating Criteria (Checklists, rubrics and scales)
- Speaking and Writing (CoWA, CoSA, STAMP, portfolio)
- Working to develop speaking and writing assessment

A View of Assessment: 30,000 ft big picture view



- Assessment Principles, Types
- Standards, Guidelines, Rating
- Speaking and Writing



Tomorrow

Reading and Listening Assessments

Backwards Design-Understanding by Design (Wiggins and McTighe, 1998)

- Integrated Performance Assessments - performance assessment units
- Topically organized assessments

Developing a performance assessment unit/topically organized unit

(Wiggins, G. & McTighe, J. (1998). *Understanding by Design*. Alexandria: Association for Supervision and curriculum Development.)

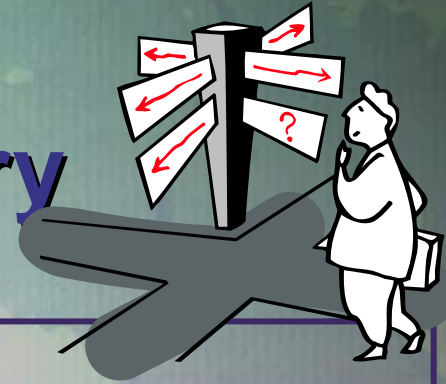
SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

The on the Ground View: Classroom and Programs



- Assessment for our purposes
- Deciding what we want students to know and be able to do
- Developing/identifying assessments aligned to goals

Assessment Inventory



Type of test/assessments	When/Why

Reasons we give assessments

Grades

Progress

Program evaluation

Instructional

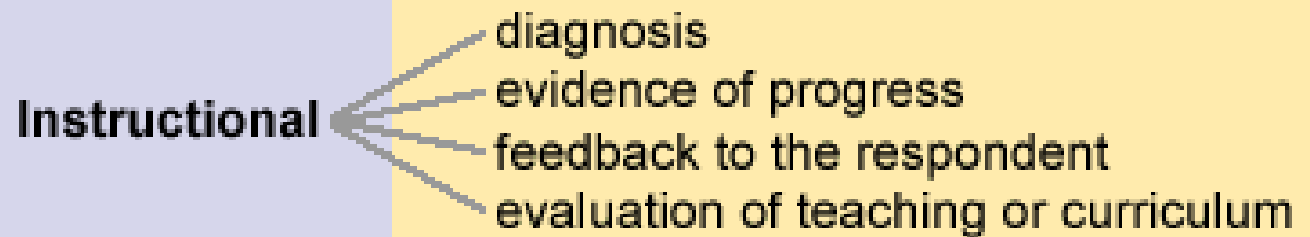
Research

The Purpose of the Assessment

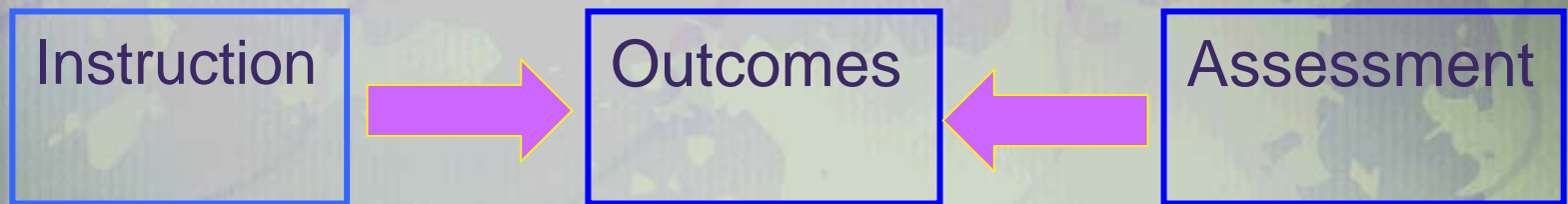
General purpose of the assessment	Specific reason for the assessment
Administrative	<ul style="list-style-type: none">general assessmentplacementexemptioncertificationpromotion
Instructional	<ul style="list-style-type: none">diagnosisevidence of progressfeedback to the respondentevaluation of teaching or curriculum
Research	<ul style="list-style-type: none">evaluationexperimentationknowledge about language learning & language use

From Cohen, A.D. (1994). *Assessing language ability in the classroom*. Boston: Heinle & Heinle, p. 23.

Focus:



Relationship among instruction, assessment and learner outcomes



Based on *Aspects of classroom-based evaluation*, Genessee & Upshur (1996, p.15)

Cohen (1994) writes that language assessment promotes "**meaningful** involvement of students with material that is central to the teaching objectives of a given course. For this meaningful involvement to take place, the goals of the assessment tasks need to reflect the goals of the course, and these goals need to be made clear to the students (p. 13)."

Cohen, A. D. (1994a). *Assessing language ability in the classroom*. 2nd Edition. Boston: Newbury House/Heinle & Heinle.

Ideally, instruction results in observable outcomes in learners, and assessment provides the means to measure how well goals have been met. Ideally, the assessment should impact instruction.

“Washback” refers to the effect that testing has on teaching.

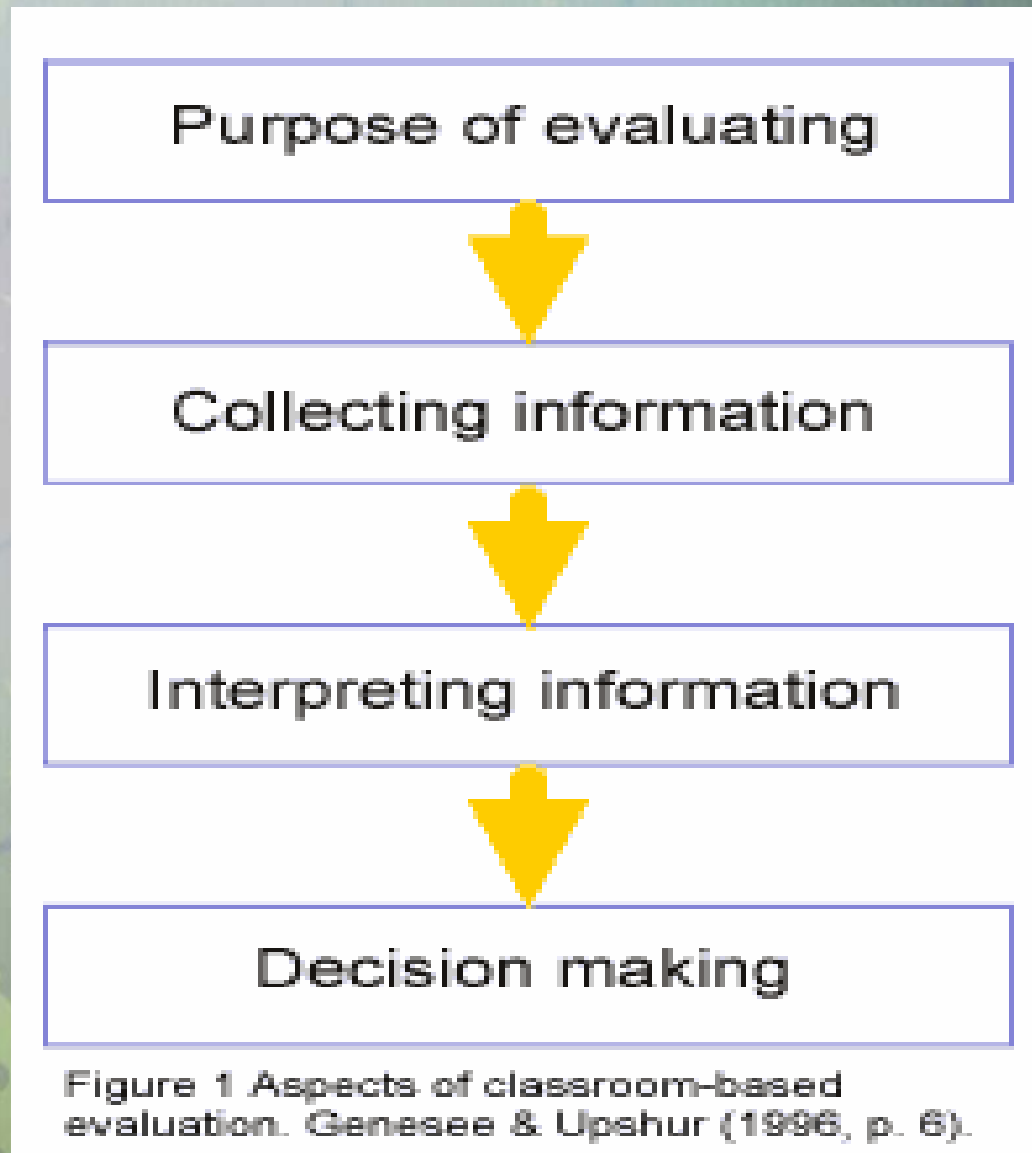
http://www.carla.umn.edu:16080/assessment/vac/WhyAssess/p_2.html

Evaluation/Measurement

a value judgment about the results of assessment data. For example, evaluation of student learning requires that educators compare student performance to standards to determine how the student measures up.

Depending on the result, decisions are made regarding whether and how to improve student performance.

Components of evaluation



Considerations in planning for instruction and assessment

Purpose of
Evaluating



Collecting
Information

Needs of



Interpreting
Information

Learners



Decision
making

Assessment - an umbrella term



Process of obtaining information that is used to make educational decisions about students, to give feedback about their progress/strengths/weaknesses, and to judge instructional effectiveness/curricular accuracy.

Terms

- ***Assessment Techniques/methods*** include tests, exhibits, interviews, surveys, observations, etc. good assessment requires a balance of techniques because each technique is limited and prone to error.
- An ***assessment instrument*** is a specific device or way to measure student learning, such as an examination, senior project, portfolio, etc.

Many ways to talk about assessment

- Purpose
- Type (techniques)
- Characteristics essential to good test
- Test taking strategies

Key Assessment Concepts

Validity

- Construct validity – test measure what it is intended to measure
- Content validity – sample is representative of target domain
- Face validity – test “looks valid” to test takers
- Washback validity – close relationship between test and teaching

- Concurrent validity – test correlates with another measure
- Predictive validity – test score predict future performance

Reliability

- Results are replicable (inter-rater reliability; internal consistency; parallel-forms reliability)

Efficiency – practicality and cost design and administration

Assessment instruments are:

- **Formative**

For improvement

- **Discrete Point**

Single set of linguistic features

- **Objective**

scoring key

- **Discrete Point**

Single set of linguistic language features

Summative

for accountability

Integrative

variety of language features simultaneously

Subjective

impression and judgment

Integrative

variety of features simultaneously

Assessment Types cont'd

- **Proficiency**

general language ability

Achievement

assess instruction

- **Norm-referenced**

compares students

Criterion-referenced

if criteria or goal met

- **Self- assessment**

students evaluate progress

Formative and Summative Assessment

Formative Assessment

Assessment for Learning

- All activities undertaken by teachers and by their students (that) provide information to be used as feedback to modify the teaching and learning activities in which they are engaged. (Black and William, 1998)
- The objective is to promote growth

Summative Assessment

Assessment of Learning

- Assessments used to determine how much students have learned as of a particular point in time
- The purpose is accountability

Based on work by R. Stiggins, Assessment Training Institute, Inc. Handout from Leadership, Learning & Assessment, LLC., MNASCD conference, March 2005

...students who understand the learning objectives and assessment criteria and have opportunities to reflect on their work show greater improvement than those who do not (Fontana & Fernandes, 1994; Frederikson & White, 1997).

Self-Assessment

- Encourages students to be self-aware of their learning
- Provides opportunity to inventory their progress
- Provides teachers with a snapshot of what students need (or their perception of needs)
- Is efficient
- Provides students with practice for peer assessment

Traditional

(Objective)

Authentic

(Alternative)

Assessments

Multiple choice

Interviews

True false

Essay

Fill in the blank

Panel discussion

Cloze

Constructed response

Traditional assessments

Quiz & Test Items - simple, content-focused questions.

- Assess factual information, concepts, and discrete skill.
- Use selected-response or short-answer formats.
- Are convergent--typically they have a single, best answer.
- May be easily scored using an answer key (or machine scoring).
- Are typically secure (not known in advance).

http://www.pgcps.pg.k12.md.us/~croom/types_of_assessment.htm

Cloze Test

- is a standard reading comprehension test.
- Originally used to assess the relative readability of written materials for school children in the United States.

Brown, J.D. (2002). Do cloze tests really work? Or, is it just an illusion? in *Second Language Studies*, 21(1), Fall 2002, pp. 79-125. Retrieved March 3, 2004, from [http://www.hawaii.edu/sls/uhwpe/sl/21\(1\)/BrownCloze.pdf#search='CLOZE%20tests'](http://www.hawaii.edu/sls/uhwpe/sl/21(1)/BrownCloze.pdf#search='CLOZE%20tests')

- Keeps the first sentence intact and then deletes every fifth word after the first sentence. (also 4th or 6th)
- Text is normally 250-300 words.
- Researchers: not a significant difference when partial answers are counted (e.g. correct part of speech)
- Passing is generally 60% or above.

Catherine E. Jones catjones@holly.colostate.edu
[http://www.cra.org/Activities/craw/dmp/awards/2003/Jones/Final_Report.p
df#search='CLOZE%20tests](http://www.cra.org/Activities/craw/dmp/awards/2003/Jones/Final_Report.pdf#search='CLOZE%20tests)

THE CLOZE TEST

- I loved the smallest lion cub best. She was the weakest, but she was the bravest of the three. I called her Elsa. If the cubs had stayed in the wild, Elsa would be dead. A lioness (2) _____ has four cubs. One dies soon (3) _____ birth, and another one is often (4) _____ weak to live long. It is (5) _____ this reason that you usually see (6) _____ two cubs with a lioness. Their (7) _____ looks after them well until they (8) _____ two years old. For the first (9) _____ she brings their food. She eats (10) _____ herself first, then she vomits it (11) _____ again.

In this way, the cubs (12) _____ manage the food.
Lions often live (13) _____ in a group called a
“pride”.
(14) _____ full-grown lions of the pride (15)
_____ all of the hunting. During their (16)
_____ year, the cubs hunt with the (17) _____,
but they are unable to kill (18) _____ their
own. They eat only what (19) _____ big lions
leave. Often very little (20) _____ for them,
so they are usually (21) _____ a very bad
condition at this (22) _____.

Fotos, Sandra S. (1991). The Cloze test as an Integrative Measure of EFL Proficiency: A Substitute for Essays on College Entrance Examination? *Language Learning*, 41:3, p. 336.

THE CLOZE TEST

- I loved the smallest lion cub best. She was the weakest, but she was the bravest of the three. I called her Elsa. If the cubs had stayed in the wild, Elsa would be dead. A lioness (2) **usually** has four cubs. One dies soon (3) **after** birth, and another one is often (4) **too** weak to live long. It is (5) **for** this reason that you usually see (6) **only** two cubs with a lioness. Their (7) **mother** looks after them well until they (8) **are** two years old. For the first (9) **year** she brings their food. She eats (10) **it** herself first, then she vomits it (11) **up** again.

In this way, the cubs (12) **can** manage the food.
Lions often live (13) **together** in a group called
a “pride”.

(14) **The** full-grown lions of the pride (15) **do** all of
the hunting. During their (16) **first** year, the
cubs hunt with the (17) **pride**, but they are
unable to kill (18) **on** their own. They eat only
what (19) **the** big lions leave. Often very little
(20) **left** for them, so they are usually (21) **in** a
very bad condition at this (22) **time**.

Fotos, Sandra S. (1991). The Cloze test as an Integrative Measure of EFL
Proficiency: A Substitute for Essays on College Entrance Examination? *Language Learning*, 41:3, p. 336.

Other assessment types

Academic Prompts - open-ended questions or problems that require the student to think critically, not just recall knowledge, and then to prepare a response, product, or performance.

They:

- Require constructed responses under school or exam conditions.
- Are open. There is not a single, best answer or a best strategy for answering or solving them
- Often are ill-structured, requiring the development of a strategy. Involve analysis, synthesis, or evaluation.

Proficiency

- “a goal for language teaching not a methodology”
- “allows teachers to take into consideration that learners may show proficiency at different levels in different modalities at any given time.”
Learning is not linear.
- teachers select authentic tasks and texts for practice and evaluation
- provides meaningful context

- does not mean that accuracy is not important
- focuses on more student activity, less teacher lecture, using vs. studying about the language, grammar taught in context, not isolation, focus on successful, rather than correct or incorrect communication
- requires judgment-based scoring based on criteria and performance standards.
- may or may not be secure.

Authentic assessments

Performance Tasks & Projects

- complex challenges that mirror the issues and problems faced by students, they are authentic (face validity).
- range in length from short-term tasks to long-term, multi-staged projects
- Feature a setting that is real or simulated: one that involves the kind of constraints, background noise, incentives, and opportunities an adult would find in a similar situation.

- Typically require the student to address an identified audience
- Are based on a specific purpose that relates to the audience.
- Allow the student greater opportunity to personalize the task.
- Are not secure. Task, criteria, and standards are known in advance and guide the student's work.

Technical Quality of Alternative Assessments

- Consequences
- Fairness
- Transfer and Generalizability
- Cognitive Complexity
- Content Quality
- Content Coverage
- Meaningfulness
- Cost and Efficiency

<http://www.ncrel.orgsdrs/areas/issues/methods/assessment/as7tech.htm>

Performance Assessments are:

- Contextualized – meaningful context for real language use
- Authentic - Students have “real” communicative purpose
- Task-based – students perform a well-defined task to elicit use of target language
- Learner-centered – tasks are realistic for students’ age, level, interests, etc.

Three key components in performance assessment

1. Tasks that are effective in eliciting the performance to be assessed.
2. Rating criteria to evaluate the quality of the performance. The criteria reflect the relative importance of various aspects of the performance, and are appropriate for the population being assessed.
3. Raters that are trained to apply the criteria and can do so consistently.

From CARLA's Virtual Assessment Center

http://www.carla.umn.edu:16080/assessment/vac/Evaluation/p_2.html
SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

Appropriate Assessments

- Imagine you are an airline passenger on a flight bound for an English-speaking country and that the pilot, copilot, and navigator are all nonnative speakers of English. As you begin your descent into a busy metropolitan airport, you may wonder about the flight crew's ability to communicate by radio in English with the ground control personnel. Which of the following situations would make you happiest?

1. Knowing that the flight crew members had all passed a multiple-choice paper-and-pencil test of listening comprehension over topics of general interest;
2. Knowing the flight crew members had all passed a face-to-face oral interview in English on general topics (such as the *ILR OPI*); or

3. Knowing the flight crew members had all passed an authentic test of oral English communication in an air-to-ground radio setting using topics based on recordings of actual conversations between air traffic controllers and airline pilots.

Evaluate the three assessments for construct and face validity, practicality and reliability.

Bailey, K.M. (1998). Learning about Language Assessment. Cambridge: Heinle & Heinle. p.209

What does research on learning say about authentic assessments?

Brain research (Jensen, 1998) and constructivist learning theory hold that knowledge is constructed by the learner and that learning occurs in context, generally a social, interactive context.

Jensen, E. (1998). *Teaching with the brain in mind*. Alexandria, VA: Association for Supervision and Curriculum Development.

Language learning increases when material is presented in meaningful content and language connects to real life purposes rather than to the mental exercise of memorizing words and rules (Curtain & Haas 1995).

Curtain H., & Haas, M. (1995). Integrating foreign language and content instruction in grades K-12. *ERIC Digest*. EDO-FL-95-07.0000.
<http://www.cal.org/resources/digest/int-for-k8.html>

The Center for Psychology in Schools and Education (American Psychological Association Board of Educational Affairs, 1997) states that successful learners are able to link new information with existing knowledge in meaningful ways and that this linkage impacts student motivation for learning.

American Psychological Association's Board of Educational Affairs. (1997). *Learner-Centered Psychological Principles: A Framework for School Redesign and Reform*. Revision prepared by a Work Group of the American Psychological Association's Board of Educational Affairs (BEA). <http://www.apa.org/ed/lcp.html>

Learning Theory

Constructivism

- Knowledge is constructed
- Learning only occurs within some context.
- Learning independent of environment is unrealistic and incomplete.
- Learner centered
- Emphasis on knowledge and conceptions students bring to the classroom

Source: Assessment by Richard Hall

- http://medialab.umr.edu/rhall/educational_psychology/2001/vl4c/assess_new.html Permission granted.

Guidelines, standards, objectives

SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

Guidelines, standards, objectives

- ACTFL proficiency Guidelines - four skills
- Language Specific Guidelines (Provisional)
- ILR (Interagency Language Roundtable)
- CEF (Common European Framework of Reference for Languages)
- National Standards for Language Learning for the 21st Century (K-12/16)



© 1999 National Standards in Foreign Language Education Project

Communication Standard

Modes of Communication

- 1.1 Interpersonal mode (spontaneous, negotiated, two-way)
Speaking, writing (e-mail, text messaging)
- 1.2 Interpretive (one-way, understand and interpret)
- 1.3 Presentational (one-way, audience, formal)
Speaking, writing

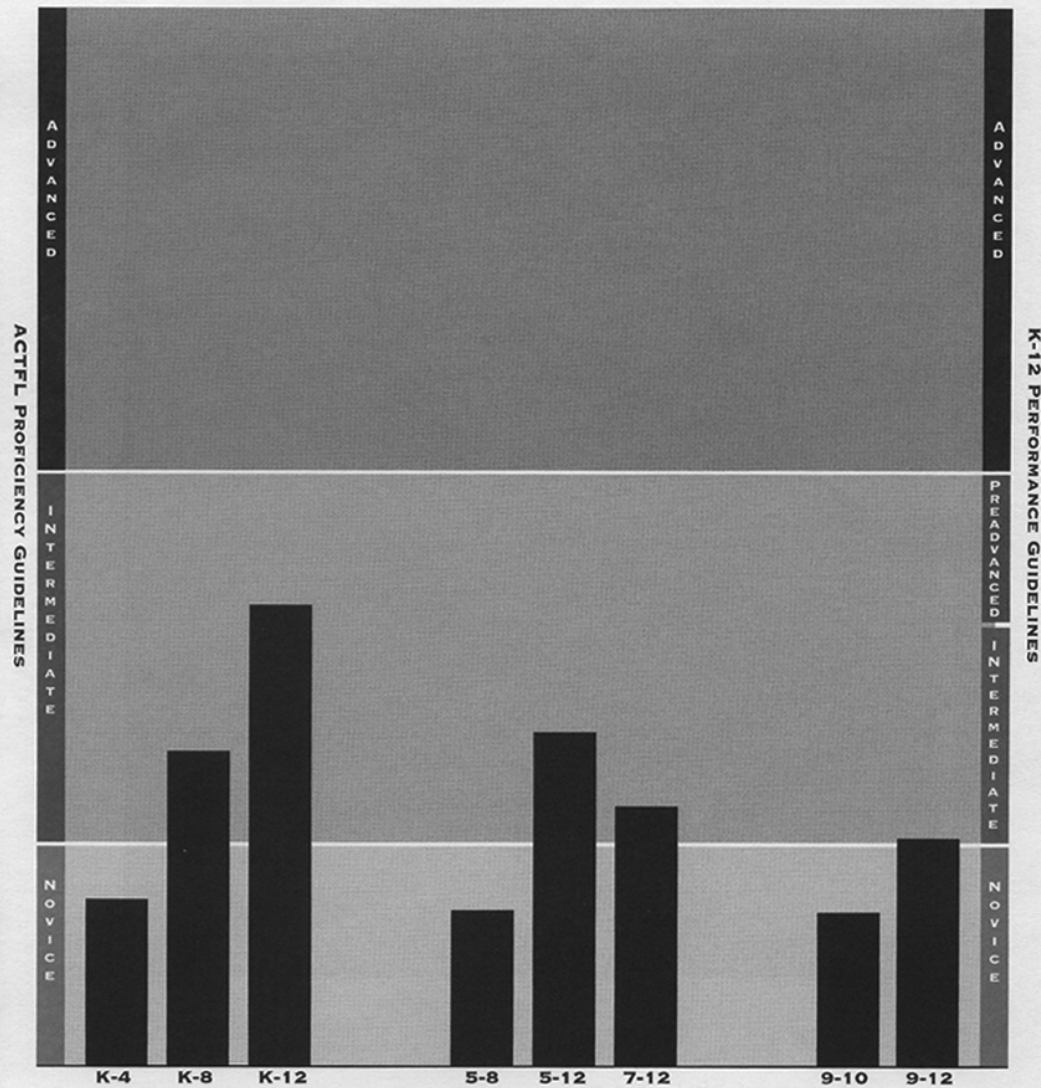


SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006



SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

Visual Representation of Anticipated Performance Outcomes As Described in the
ACTFL Performance Guidelines for K-12 Learners¹



¹ Descriptors are based on information gathered from foreign language professionals representing a variety of program models and articulation sequences. Descriptors are appropriate for languages more commonly taught in the United States. Descriptors assume a sustained sequence of Standards-based, performance-outcome, language instruction.

Postsecondary levels

University of Minnesota (equivalent of two years)

- IM/writing, IH reading and listening for F,G,S,I
- I/M reading and listening and IL for writing and speaking for Japanese, Finnish, Hindi, Korean

First Year Swahili Language Objectives

In order to demonstrate knowledge of first-year **Swahili**, students must:

- initiate and understand simple dialogues;
- handle basic social topics and functions associated with each;
- express preferences, likes, and dislikes on simple topics;
- offer descriptions;
- be aware of culturally appropriate verbal and non-verbal gestures and expressions;
- have encountered and used all constructions of Swahili grammar needed to accomplish these topics and functions.

<http://language.stanford.edu/requirement/swahili.html>

First Year Hebrew Language Objectives

In order to demonstrate knowledge of first-year **Hebrew** (Modern Israeli) at Stanford, students must:

- have acquired an understanding of the basic structures of Hebrew;
- must have a working knowledge of approximately 1000 to 1200 words;
- be able to express themselves both in writing and speaking, informally, on topics in the domains of study and work, family and friends, and daily routines;
- be able to read unvocalized texts of a higher and more formal level related to Israel, Jerusalem, Judaism, and on certain academic topics.

<http://language.stanford.edu/requirement/hebrew.html>

First Year Arabic Language Objectives

In order to demonstrate knowledge of First Year **Arabic** at Stanford, students must:

- demonstrate command of the Arabic script in writing and reading;
- have worked in writing and speaking with all basic grammatical constructions;
- have at least a rating of Novice-High in listening, speaking, reading, and writing on the ACTFL-FSI scale or its equivalent.

<http://language.stanford.edu/requirement/arabic.html>

Stanford Proficiency Notation

- The notation "Proficiency in (Language)" appears on the official transcripts of students whose levels of achievement are equivalent to those that an excellent student is expected to demonstrate late in the third quarter of the third year of study in that language.
- In order to receive the notation, the student must be given a rating of Advanced on the Foreign Service Institute/American Council on the Teaching of Foreign Languages (FSI/ACTFL) oral proficiency scale, except in the Asian Languages and Russian which require an Intermediate High rating.

<http://language.stanford.edu/requirement/html>

ACTFL OPI* and College Credit

American College of Education's (ACE) College Credit Recommendation Service (CREDIT)

- connects workplace learning with colleges and universities by
- adults gain access to academic credit for formal courses and examinations taken outside traditional degrees

<http://www.actfl.org/i4a/pages/index.cfm?pageid=3642#special>

*Oral Proficiency Interview

- Albanian, Arabic, Cambodian, Cantonese, Croatian, Czech, Dutch, Egyptian, English, Farsi, Flemish, French, German, Greek, Haitian Creole, Hebrew, Hindi, Hmong, Indonesian, Italian, Japanese, Khmer, Korean, Lao, Malay, Mandarin Chinese, Norwegian, Polish, Portuguese, Punjabi, Russian, Serbian, Slovak, Spanish, Swahili, Tagalog, Thai, Ukrainian, Urdu, and Vietnamese.

ACTFL Oral Proficiency Interview (OPI)

“is a valid and reliable means of assessing how well a person speaks a language. It is a 20-30 minute face-to-face or telephonic interview between a certified ACTFL tester and an examinee. The interview is interactive and continuously adapts to the interests and abilities of the speaker. The speaker’s performance is compared to the criteria outlined in the ACTFL Proficiency Guidelines - Speaking (Revised 1999). The interview is double rated and an Official ACTFL Oral proficiency Certificate stating the candidate’s proficiency level is issued to the candidate.”

ACTFL. Retrieved October 9, 2006 from <http://www.actfl.org/i4a/pages/index.cfm?pageid=3642#profassess>

Summary of ACE Credit Recommendations for Official ACTFL OPI Ratings

	Category I	Category II	Category III	Category IV
Official OPI Rating	<i>Dutch, English* , French, Haitian Creole, Italian, Norwegian, Portuguese, Spanish, Swahili and Swedish.</i>	<i>German, Modern Greek, Hindi, Indonesian, Punjabi and Urdu.</i>	<i>Cambodian, Czech, Hmong, Hebrew, Hungarian, Lao, Polish, Russian, Serbo-Croatian**, Slovak, Tagalog, Turkish, Thai, Ukrainian and Vietnamese.</i>	<i>Arabic, Cantonese, Japanese, Korean and Mandarin.</i>
Novice High/ Intermediate Low	2 LD	2 LD	3 LD	3 LD
Intermediate Mid	4 LD	5 LD	6 LD	6 LD
Intermediate High/ Advanced Low	8 LD	8 LD + 2 UD	6 LD + 3 UD	6 LD + 3 UD
Advanced Mid	10 LD	8 LD + 4 UD	6 LD + 6 UD	6 LD + 6 UD
Advanced High/ Superior	10 LD + 2 UD	8 LD + 4 UD	6 LD + 6 UD	6 LD + 6 UD

Legend

*English is treated as a foreign language.

** Serbian and Croatian have been combined to Serbo-Croatian.

Credit recommendations are based on a semester hour.

LD = Lower division baccalaureate/associate degree category.

UD = Upper division baccalaureate degree category.

The 10 ACTFL OPI language proficiency ratings are: Superior, Advanced High, Advanced Mid, Advanced Low, Intermediate High, Intermediate Mid, Intermediate Low, Novice High, Novice Mid, Novice Low.

Expected Benchmarks

Year 1 Novice High?

Year 2 Novice High/Intermediate-Low?

Year 3 Intermediate-Low/Mid?

Year 4 Intermediate-Mid?

Test Frameworks

Rating scales and rubrics

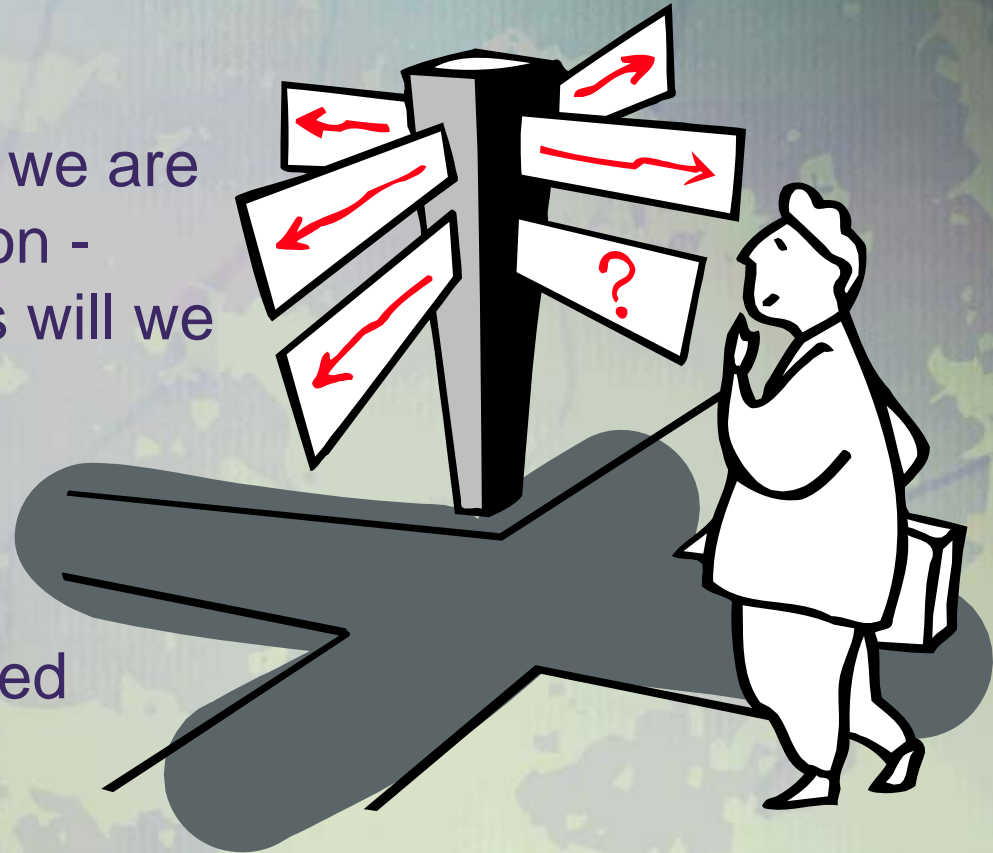
SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

Destination: where are we going?

Roadmap

How will we know we are at our destination - what landmarks will we look for ?

What will we need to get there?



Some Assessment Frameworks

- OPI-Oral Proficiency Interview (ACTFL)
- SoPI-Simulated Oral Proficiency Interview
tape-mediated (Center for Applied Linguistics)
- CLEAR Test Development Workbook and
Video (Center for Language Education and
Research)

- CoSA and CoWA (Contextualized Speaking/Writing Assessments)
- Backwards Design-approach to test/task design
- Integrated Performance Assessments

CLEAR Test Development Workbook and Video

Unit 1: Rules for Item Writing

Unit 2: Test Purposes

Unit 3: Test Item Grid

Unit 4: Detailed Text-Based Item Design Tables

Unit 5: Detailed Oral/Aural Item Design Tables

Unit 6: Test Specifications

Unit 7: The Test Development Process and Test
Administration

Center for Language Education and Research. (1998). *Test Development Workbook and Video*. East Lansing, MI: Michigan State University Board of Trustees.

<http://clear.msu.edu/teaching/testvid/index.html>

Before designing a test, there are at least five questions to consider.*

- 1. What are the instructional goals and objectives?
aligning course objective with assessment
- 2. What type of syllabus design is being used?
notional-functional
situational
skill-based
task-based
topical
combination

- *Center for Language Education and Research. (1998). *Test Development Workbook and Video*. East Lansing, MI: Michigan State University Board of Trustees.

3. What is the purpose of the test?

proficiency
placement
diagnostic
achievement

4. How will this test fit into the curriculum, class or program?

5. How much time is required and available to create, administer, score, and make decisions based on this test? (Practicality)

- * Center for Language Education and Research. (1998). *Test Development Workbook and Video*. East Lansing, MI: Michigan State University Board of Trustees. p. 1.

SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

Backwards Design

- “Backward design may be thought of as purposeful task analysis: Given a task to be accomplished, how do we get there? Or, one might call it planned coaching: What kinds of lessons and practices are needed to master key performances?...Rather than creating assessments near the conclusion of a unit of study (or relying on the tests provided by textbook publishers, which may not completely or appropriately assess our standards), backward design calls for us to operationalize our goals or standards in terms of assessment evidence as we begin to plan a unit or course.”

Wiggins & McTighe (1998).

Rating Methods

Checklists

Rubrics

Scales

Rating methods

Objective and subjective evaluation

If we view objectivity and subjectivity of evaluation along a continuum, we can represent various assessment and scoring methods along its length.



Rating methods

- Objectively – traditional assessments

Pros

Cons

- Subjectively – authentic assessments

Pros

Cons

Rater reliability

In assessment, we strive to ensure two types of reliability:

inter-rater (raters agree with each other) and

intra-rater (a rater gives the same score to a performance rated on separate occasions)

Checklists, Rubrics and Scales

Use **checklists** to:

- Record observed performance.
- Keep track of progress over time, e.g., inventory of skills at beginning and end of courses.

Receptive oral skills – sample checklist

- Understands simple directions.
- Understands simple sentences.
- Understands simple yes/no questions.
- Understands vocabulary appropriate to age.
- Understands meaning of different intonation patterns.
- Understands more complex directions.
- Understands rapid speech.
- Understands language in classroom situation.
- Understands language of peers.

<http://www.carla.umn.edu/assessment/vac>

Adapted from Genesee, F. & Upshur, J.A. (1996). *Classroom-based evaluation in second-language education*. Cambridge: Cambridge University Press, p. 88.

Checklists

- Less informative than scaled rubrics
- Efficient
- Record observed performance
- Use simplifies rubric construction by specifying non-negotiables (Donna Clementi, 2002)

Advantages and Disadvantages

Advantages

- Easy to construct and use.
- Align closely with tasks.
- Effective for self and peer assessment.
- Make learners aware of task requirements, allowing them to self-monitor progress.
- Useful for sharing information with students and other stakeholders.

- [\(Brindley, 1989; Genesee & Upshur, 1996; Tedick, 2002; Underhill, 1987\)](#)

Disadvantages

- Provide limited information about how to improve performance.
- Do not indicate relative quality of performance.

Rubrics

- rating scales used with performance assessments
- increasingly used to evaluate performance assessments.
- provide an indication of **quality** of performance/student work
- Provide detailed feedback to learners

Rubrics & Scales

- Scales and rubrics differ in their specificity: scales generally can be applied to more than one task; rubrics are more closely aligned with particular tasks.
- Holistic rubrics and scales are used to assign one integrated evaluation to a performance, e.g., *ACTFL Proficiency Guidelines*.
- Analytic rubrics and scales are used to evaluate the various dimensions of a performance.

Why use rubrics?

- Set anchor points along a quality continuum rather than *right* or *wrong*
- Increase construct validity
- Align assessment to curriculum and instruction
- Focus on the most salient goals
- Expectations are clearer to students (and to yourself)

- Provide specific feedback to students
- Well-designed rubrics increase assessment reliability by setting criteria that can be consistently applied by raters
- The Virtual Assessment Center at CARLA provides an extensive tutorial on rubrics

http://www.carla.umn.edu/assessment/vac/Evaluation/p_4.html

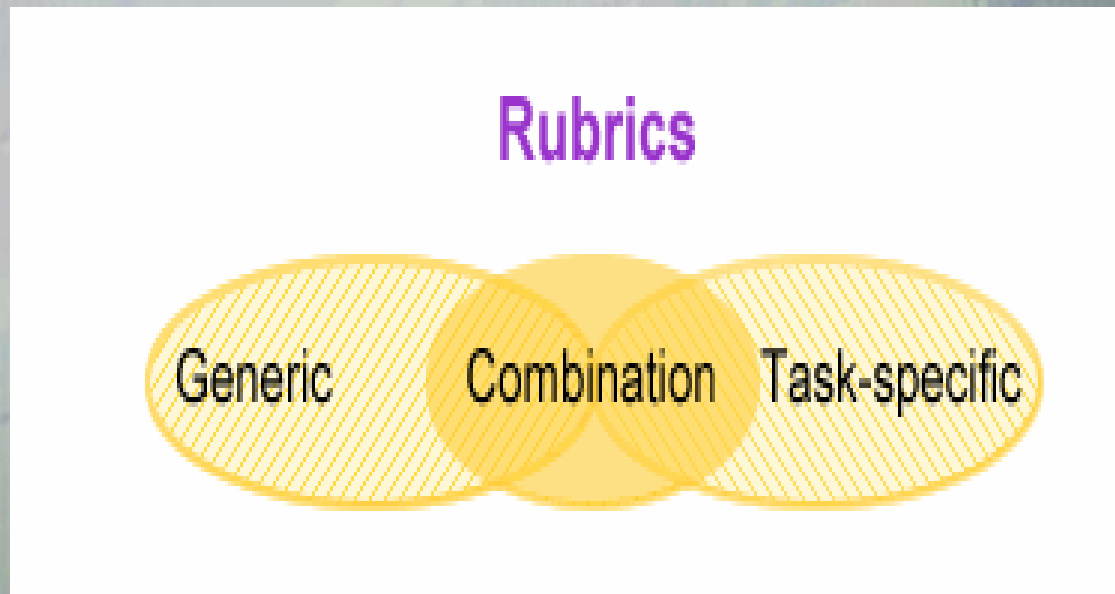
Rubrics should be:

- Clear - define excellent, good or work needing improvement
- Democratic – consistent, consistently reinforced, understood by students that grading is not subjective
- Involving – students should be involved in creating rubrics
- Engaging – allow students to evaluate own work

From Blaz, C. (2001). *A collection of performance assessment tasks and rubrics*.

Eye on Education: Larchmont, NY.

Types of Rubrics



Compare:

ACTFL Performance Guidelines for K-12 Learners

- Comprehensibility
- Comprehension Language
- Control Vocabulary Cultural
- Awareness Communication Strategies

To:

Analytic Writing Scale for the Spanish FLIP Program, University of Minnesota*

Content (30 pts.)

Organization (20 pts.)

Language use/Grammar/ Morphology (25 pts.)

Vocabulary/Word usage (20 pts.)

Mechanics (5 pts.)

Multitrait Rubric for Recipe Presentation

- Organization and Clarity
- Fluency and Pronunciation
- Use of the Imperative (or other form)

Primary trait rubrics

- Performance is scored on the main criterion for success on the task.
- Task*: Write a persuasive letter to the editor of the school newspaper.

Primary Trait: Persuading an audience*

0	Fails to persuade the audience
1	Attempts to persuade but does not provide sufficient support.
2	Presents a somewhat persuasive argument but without consistent development and support
3	Develops a persuasive argument that is well developed and supported.
<p>* Tasks/rubrics can be found in Minnesota Articulation Project. (2002). <i>Proficiency-oriented language instruction and assessment: A curriculum handbook for teachers (Rev Ed.)</i>. CARLA Working Paper Series. D. J. Tedick (Ed.). Minneapolis, MN: University of Minnesota, The Center for Advanced Research on Language Acquisition.</p>	

Creating a Rubric

Step 1: Generate potential dimensions

- Rank order the potential dimensions from most to least important; eliminate the "non-negotiables".

Step 2: Select a reasonable number of dimensions

3-7?, 4-8?

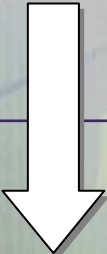

- "How many is enough? ... There's no one correct answer, but it might help if you consider your purpose for this measurement. If it's diagnostic and formative, err on the side of more dimensions rather than fewer.
- If you just want to be able to give a summative evaluation of your students' performance for this particular lesson, fewer dimensions are OK." (Triton/Patterns Summer Symposium. (1999). *Creating A Rubric for a Given Task*. San Diego City Schools.

Available online at <http://projects.edtech.sandi.net/staffdev/tpss99/rubrics/rubrics.html>

Dimensions

Dimensions	Accomplished <i>Exceeds expectations</i> 3 <i>Excellent</i>	Developing <i>Meets expectations</i> 2 <i>Average</i>	Beginning <i>Not there yet</i> 1 <i>Needs work</i>
			
SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006			

Write Benchmark descriptors

	Accomplished <i>Exceeds expectations</i> 3 <i>Excellent</i>	Developing <i>Meets expectations</i> 2 <i>Average</i>	Beginning <i>Not there yet</i> 1 <i>Needs work</i>	
Dimensions				
	 Benchmark Descriptions			

Avoid:

- "Squishy" descriptors. For example, these descriptors were used to evaluate essays:

3 shows depth	2 lacks depth	1 no depth
---------------------	---------------------	------------------

5 well- balanced	4 moderately well-balanced	3 not so well- balanced	2 lack of balance	1 total lack of balance
------------------------	----------------------------------	----------------------------------	-------------------------	-------------------------------

Unnecessary negative language:

3 Creative introduction	2 Adequate introduction	1 Boring introduction
-------------------------------	-------------------------------	-----------------------------

Evaluating an ESL speaking assessment

According to the number on your sheet, in groups of four or five:

- please read the purpose for which you will use this speaking assessment
- Read the ESL speaking assessment on the handout
- make a rubric with criteria that meet the assigned purpose
(to consider level, consult your handouts if needed)
- Listen to the student, and rate the sample using the rubric your group produced.

English Speaking CoSA

Meeting your new tandem partner

It's the second week of school, and the ESL program has arranged a reception for international students to meet their tandem partners. You are paired up with Paula, who is from southern Iowa, and studies linguistics.

Task 1

Paula wonders about you: where you're from, how old you are, what you're studying, and so on. Wait until you hear Paula's question, then **tell her about yourself.**

Task 2

Paula asks you about your classes: she's curious about the ESL program. You tell her which classes you're taking and your opinion of each class. You also tell her a few things that you do in your classes. Wait until you hear Paula's question, then **describe your classes to her, and include the activities students do in each class.**

Task 3

You and Paula begin talking about dating in the United States. She wants to know how you meet people in your home country, and what dating customs are like there: are parents involved, do people go out in groups, and so on. Wait until you hear Paula's question, **then tell her about dating in your country.**

Task 4

Later that day, you run into Paula in the library—you're both tired of studying, and you start to talk about families. She tells you all about her family in Iowa, and then wants to know about yours: your sisters, brothers, parents, etc.: their ages, where they work, what they do in their spare time, and so on. Wait until you hear Paula's question, then **describe your family to her.**

Task 5

You would like to learn more about Paula and her life in Iowa: her friends, her interests, the town she lives in, etc. Wait until you hear Paula speak, then **ask her at least five questions to find out more about her life in Iowa.**

Task 6

You start thinking ahead to the weekend, and Paula wonders what there is to do here at the University of Minnesota and in the Twin Cities. She wants to get to know the area and make some new friends. Wait until you hear Paula's question, then **make at least 4 suggestions as to what she might do this weekend.**

Task 7

You and Paula decide that you will meet later for lunch to make plans. Wait until you hear Paula's question, then **tell her where you go to eat lunch, what you like to eat, and that you will see her later.**

Speaking Assessment Scoring Criteria - Intermediate-Low

Task Fulfillment

The speaker demonstrates the ability to:

Provide responses of sufficient length and substance for the evaluator to apply the scoring criteria.

Speak about the assigned topics and perform the assigned functions

Ask and answer questions.

Vocabulary

The speaker:

Demonstrates an adequately broad vocabulary to address the topics.

Does not have to resort to excessive repetition of words or phrases.

Communicates effectively in the target language despite errors in word choice and occasional lapses into English.

(i.e., errors and lapses do not substantially obscure the message.)

Discourse

The speaker demonstrates the ability to:

Create with language.

Use basic sentence-level discourse.

Go beyond memorized chunks and patterns.

Use more than one sentence pattern.

Communicate effectively in the target language despite occasional direct translations from English, which are characteristic of this level.

Present Time

The speaker:

Can express present time (i.e., there are few errors when conjugating common verbs in the present tense).

May express the immediate future (*to go* + infinitive), and/or use adverbs of time to express future actions.

Errors in grammar, pronunciation, and vocabulary do not substantially obscure the speaker's intended meaning for a sympathetic listener.

Accuracy

Errors in grammar, pronunciation, and vocabulary do not substantially obscure the speaker's intended meaning for a sympathetic listener.

Assessing for Proficiency

Writing and Speaking

SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

Proficiency assessment

- Purpose: What do I want to know from my assessment?
- Method: How will I assess, and is my assessment method appropriate?
- Audience: Who am I assessing
- Validation: How will I know if I assessed what I intended to assess?
- Interpretation: How will I evaluate performance and what will results/ scores mean?
- Use: How will I use the results/scores?
- Reliability: How will I know that my measurement instrument is accurate?
- Practicality: Is my measurement instrument practical, given my resources?

SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

Assessing Writing and Speaking

Writing: The good news

- Purpose of writing assessment: demonstrate proficiency in the target language
- Easiest to do
- form most frequently used by textbook companies
- good test topics/ideas available from other subject areas

From Blaz, D.(2001). *A collection of performance assessment tasks and rubrics*.
Larchmont: Eye on Education. p. 59-62

.....and the bad news

- We have difficulty deciding on how to grade them
- Rough drafts to assess or not assess
- use a rubric like that for the final product
- incorporate the draft as a category on the rubric
- targeted peer edit-prevents “help” from outside
- need to include students’ improvement/progress on the rubric to reinforce improvement

Concerns in Assessing Oral Language

- What to assess: content or form
- How to assess: objective or subjective
- When to assess
- Logistics: practical concerns – numbers of students, administration and rater time required

Considerations

- Purpose
- Type
- Level
- Validity (content and face)
- Reliability
- Practicality

Possible oral assessment tasks

- Anecdote, dramatization, monologue, ballad, explanation, narration, book report, fairy tale, newscast, campaign speech, free verse, oath, character portrayal, interview, recipe, choral reading/speech, jingle/ad, riddle, debate, demonstration, joke, role-play, dialogue, lecture, seminar, discussion, lesson, skit, documentary, mock interview, weather report.

*Blaz, D. (2001). *A collection of performance tasks and rubrics: Foreign languages*. Larchmont, NY: Eye on Education, p.35.

Possible writing assessment tasks

- Guidebook, pamphlet, autobiography, handout, parody, biography, headline, pen pal letter, book report, interview script, petition, booklet/brochure, job description, play, poem, business letter, joke, poster, cartoon, journal, prediction, celebrity profile, law, puppet show, checklist, lesson plan, quiz, comic book, list, recipe, commercial script, log, report, comparison, lyrics.
- *Blaz, D. (2001). *A collection of performance tasks and rubrics: Foreign languages*. Larchmont, NY: Eye on Education, p.59.

Cont'd.

- Review, magazine/news article, riddle, creative writing, critique, metaphor, script, description, movie review, short story, dialogue, newsletter, slogan, diary/journal, new story ending, story problem, editorial, essay, notes, survey, fact, file, oath, telegram, fairy tale/myth, observation sheet, travel log, glossary, outline, yearbook entry, advertisement
 - *Blaz, D. (2001). *A collection of performance tasks and rubrics: Foreign languages*. Larchmont, NY: Eye on Education, p.59.

STAMP

- The STAMP test is realia-based and Computer Adaptive. It measures interpretational and presentational modes including reading, writing and speaking proficiencies.
- Benchmarks are based on the ACTFL Proficiency Guidelines and measure from Novice-Low through Intermediate-Mid.
- STAMP is age-appropriate for grades 7 through 16.

LLS. Retrieved October 1, 2006 from
<http://www.stamptest.net/stamp0405/stamptest/>



<http://www.stamptest.net/stamp0405/stamptest/>

SALRC Workshop, University of California at Berkeley, Oct. 27-28, 2006

STAMP

Standards-based Measurement of Proficiency

Estampas heretempes

日本語

(Japanese)

Español

(Spanish)

Français

(French)

Deutsch

(German)

简体中文

(Simplified Chinese)

繁體中文

(Traditional Chinese)

Login To The Website Below:

Test Code:

Password:

Your Full Name:

(e.g., John Smith)

Attention students!

Please make sure you typed your name correctly. If you need to resume the test for any reason, you will need to re-type it exactly as it appears here.



A Service of Language Learning Solutions www.onlinells.com

© 2006, Language Learning Solutions

El Siglo de Oro

Llamamos Siglo de Oro a un tiempo en el que España fue la primera potencia mundial. España descubría, exploraba y conquistaba nuevas tierras en otros continentes; los soldados españoles vencían en los campos de batalla de Europa; los galeones (barcos) traían oro de América del norte y sur; la moneda española era la más valorada del mundo; la vestimenta y las costumbres españolas estaban de moda. Además, en ese período florecieron, como nunca antes en España, el arte (pintura, escultura y arquitectura) y la literatura (poesía, novela y teatro.)



According to the text above, which of the following statements about the Spanish Golden Age is NOT true?

- Spain prospered economically.
- Spanish art and literature were considered unimportant.
- Spaniards explored many other continents.
- Spain was an important world power.

Submit

Continue

[Stop Test](#)

STAMP

Standards-based Measurement of Proficiency

Estampage Sample

Your Swiss host brother, Pierre, is shopping online for some new dress clothes. He would like your opinion before he finalizes his purchase:



The screenshot shows a web browser window with the address bar displaying "@ alamode.com". The page content includes a prompt "Cliquez sur l'image pour l'agrandir" above a photograph of a man in a light-colored suit and patterned tie. To the right of the image is a product description for "LE COSTUME CLASSIC", detailing its features: "Veste entièrement doublée. Deux poches passepoilées et une poche paysanne. Poignets quatre boutons. Pantalon monté sur ceinture, fermeture patte bouton et crochet, deux poches italiennes. A l'arrière, deux poches passepoilées à boutons. 100 % laine, doublure : 100 % polyester". Below the description, the price is listed as "prix: 199,95 €". A color selection section labeled "couleur:" has three radio buttons: "gris clair" (selected), "bleu marine", and "beige". At the bottom of the product area is a yellow button labeled "Ajoutez au panier". A blue link "Fermez cette fenêtre" is located below the image.

The item of clothing advertised on this Web page is:

- a suit
- a tie
- a jacket
- a dress shirt

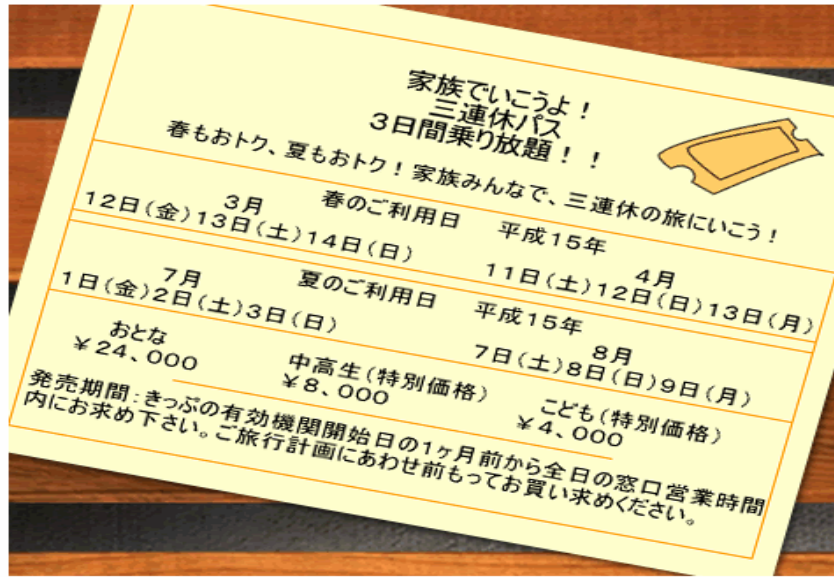
Submit

STAMP

Standards-based Measurement of Proficiency

Sample

On his way to the university this morning, Yoshi saw the following flier at the bus stop.



What is this flier advertising?

- concert tickets
- theater tickets
- plane tickets
- train tickets

Submit

[Stop Test](#)

Portfolio Assessment

- A purposeful collection of student work that exhibits the student's efforts, progress and achievements in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit and evidence of student self-reflection."

(Paulson, Paulson, Meyer 1991) Judy Kemp at:

<http://www.anglit.net/main/portfolio/default.html>

- systematic longitudinal collection of student work created in response to specific, known instructional objectives and evaluated in relation to the same predetermined scoring criteria

Portfolio Assessment in the Second Language Classroom (1998). National Capitol Language Resource Center)

Advantages of Assessment Portfolios

- Authentic picture of learning
- Improved teaching and student learning
- Electronic portfolios are easily managed
- Electronic Portfolios: Students, Teachers, and Life Long Learners

<http://www.eduscapes.com/tap/topic82.htm>

Challenges of Assessment Portfolios

Lower comparability and reliability

- do not easily translate into grades

Cost

- ongoing expense if used for large scale testing
- aligning the assessment tasks with curriculum and developing the scoring criteria and scoring tools.

Scoring

- developing and using scoring criteria requires labor intensive staff discussion and training.
- scoring student work using predetermined criteria is more time consuming than scoring a single-occasion, norm-referenced test; requires intensive staff development
- Fortunately, these same activities also lead to improved teaching and learning.

Portfolio Assessment in the Second Language Classroom (1998). National Capitol Language Resource Center)

Design Steps in Foreign Language Portfolio Assessment

- Set Assessment Purpose
- Identify Instructional Objectives
- Match Tasks to Objectives
- Set Criteria
- Determine Organization
- Monitor Progress
- Evaluate the Portfolio Process

- *Portfolio Assessment in the Second Language Classroom* (1998). National Capitol Language Resource Center) p. 14.



Portfolio Assessment in the Foreign Language Classroom

[About Us](#) [Modules](#) [Forms](#) [Resources](#) [Contact Us](#) [NCLRC](#)

[PA Home](#) > [Modules Main Page](#) > [Part 3: How to Use this Manual](#) > **How to Use This Manual**

How to Use this Manual

[Go back to Part 2](#)

[Go on to part 4](#)

Think big, plan small !!!

- [3:1 - Starting out](#)
- [3:2 - Cooperation](#)
- [3:3 - Adapting materials](#)
- [3:4 - Language of instruction](#)
- [3:5 - Using the Teacher Self-Efficacy Questionnaire](#)

Forms in this section:

[Teacher Self-Efficacy Questionnaire](#)

[Selecting a portfolio class and support system](#)

<http://www.anglit.net/main/portfolio/default.html>

CoSA and CoWA (Contextualized Speaking/Writing Assessments)

Proficiency-based assessments

- End of course of study; e.g., end of a level of language, end of semester
- Development template/grid/framework adapts to level and appropriate topics and stimuli
- Miniguide for Developing Speaking and Writing Tasks for Second Language Assessment available at:

<http://www.carla.umn.edu:16080//assessment/MLPA/pdfs/miniguide.pdf>

Minnesota Language Proficiency Assessments
Contextualized Speaking Assessment (CoSA) Development Worksheet: Level: Intermediate (Part 1)

I. CoSA Theme	Segments	II. Topic	III. Situation
Provide students with rich description of the general setting in which they are asked to imagine themselves speaking in L2. ¶ ¶ ¶ • What is the general setting? What motivates the speaker to speak L2?	<input type="checkbox"/>	Hobbies Personal possessions Leisure time Clothing Celebrations Rooms/housing Family and friends Daily activities Seasons Other areas pertaining to self & immediate environment	Who is the speaker? (speaker represents self) To whom does s/he speak? (age, number, status of audience) Where does conversation occur? When does conversation occur? Why does conversation occur? <input type="checkbox"/>
<input type="checkbox"/>	Warm-up <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Is the speaker in a plausible situation where he/she can respond with personal experiences? yes no <input type="checkbox"/>	Segment 1 <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Segment 2 <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• With how many characters does the speaker interact? Who are the characters (age, status, occupation)? <input type="checkbox"/>	Segment 3 <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Segment 4 <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Segment 5 <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	Wind-down <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Segments	IV. Functions	V. Task	VI. Prompt
<ul style="list-style-type: none"> □ ¶ Asking questions¶ ¶ Making suggestions¶ ¶ Describing in areas that pertain to self and immediate environment¶ ¶ Narrating based on a sequence of pictures¶ ¶ Express likes and dislikes¶ ¶ □ 		<ul style="list-style-type: none"> ¶ Specify the speaker's task □ 	<ul style="list-style-type: none"> ¶ Write the prompt in the target language. □
Warm-up¶ □	□	□	□
Segment 1¶ □	□	□	□
Segment 2¶ □	□	□	□
Segment 3¶ □	□	□	□
Segment 4¶ □	□	□	□
Segment 5¶ □	□	□	□
Wind-down □	□	□	□

Assessing Listening

- Least understood
- One of most important areas of language assessment
 - Washback into classroom practice
 - Student focus on listening skills/strategies
- Central to teaching and assessing language proficiency

- Closely linked to speaking
- Students/teachers are required to engage in speaking activity
- Should be tested when speaking cannot be

Source: J. Charles Alderson, Lyle F. Bachman in Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

The Listening Construct

- Complex skill i.e. involves multiple types of knowledge
- Syntax, semantics, discourse
- Culture (formal and informal situations)
- Non-linguistic knowledge about the way the world works

Metaphor for listening:

Running on a riverbank and speaking to someone in a boat. You have to keep running so the person won't get out of range.

You can't stop the boat for words you don't understand and don't have time to translate everything into English

because you will slow down and can't keep up with the boat.

- Source: Blaz, D. (2001). *A collection of performance tasks and rubrics*. Larchmont, New York: Eye on Education. p. 91.

Listening Assessments

Determine purpose

- Academic or real-life
- Select a stimulus

authentic text, video, audio, monologue,
realistic text, dialogue

- Determine appropriate response format
academic responses
realistic responses

Academic responses

- Discrete-point exercise
- Complete-open ended statements
- Dictation
- Cloze
- Content question/answer
- Think aloud/retrospection

Realistic-responses

- Oral response
- Main idea
- Paraphrase
- Summarize/condense
- Transfer to another skill (message, notes)
- Converse

From: Long, D.R. and Macián, J.L. (1994) Listening Skills and Assessment. In C. Hancock (Ed.), *Teaching, Testing, and Assessment: Making the Connection*. Lincolnwood, IL: National Textbook. (p.130).

Test Characteristics: French CoLA, Intermediate-Low Level

Context/Theme	Text Type(s)	Topic(s)
Returning from a trip	Informal conversation among peers; PA announcement	Travel, transportation, airport information, vacation plans, invitations, instructions
Café; restaurant	Informal and formal conversations	Travel, leisure, work, friends, money, ordering food and beverages, likes and dislikes
Making plans for the day	Informal conversations	Leisure, transportation, food, film, invitations, preferences
Doing errands	Informal and formal conversations, radio broadcast	Directions, news event, music, housing, post office transactions, money, appointments, instructions
Friends and family	Informal conversation	Greetings, introductions, studies, leisure, family, health, sports
Shopping	Informal conversations, news report, announcement, commercial	Money, clothing, colors, product description, leisure, preferences, current events



© 2001 Regents of the University of Minnesota

Listen to the Baking bread from

<http://dev->

[carla.umn.edu/assessment/VIB/english/items_listening/bakingbreadmain.htm](http://dev-carla.umn.edu/assessment/VIB/english/items_listening/bakingbreadmain.htm)

What other ways could you assess the segment?

Assessing Reading

Reading Comprehension

Three-level process description developed by Herber. (Thrasher, 2000)

- Literal - Reader examines words, what they say and information presented.
- Interpretive - Reader looks for relationships in statements and meaning conveyed. Reader's experience with the topic influenced by previous knowledge.

Applied - Reader takes literal (what author said) and interpretive (what author meant) applies to previous knowledge on subject to deepen understanding. (Herber, 40)

Reading Assessment

Intermediate Level Specifications Planning Guide

Content/Topic	Function/Task	understand main ideas	understand main gist	understand some supporting detail
narratives on familiar topics				
TV/radio program schedules				
forms (travel, bank, postal)				
menus				
messages and memos				
simple letters and postcards				
ads and labels				
simple instructions				
newspaper headlines				
tables of contents				
maps				
personal biographical information				
restaurant/foods				
asking/giving directions				
activities/hobbies				
transportation				
shopping/making purchases				
lodging				

Traditional assessments

- Matching,
- Cloze
- Fill in the blank
- Short answers

Difficult to assess with performance assessment

- Mental activity
- Assessed with spoken or a written product
- Method depends on the skill we wish to assess

Blaz, D. (2001). *A Collection of Performance Assessment Tasks and Rubrics*.
Larchmont, NY: Eye on Education.

Reading Activities

Organize

Job application

Mind Map

Out line

Timeline

Travel log

Venn Diagram

Demonstrate

Ad

Critique

Mock Interview

Process

Discussion

Journal

Retelling

Script

Seminar

Create

Letter

New ending

Updated Version

Video

Putting the skills together

Integrated performance assessments/topically organized assessments

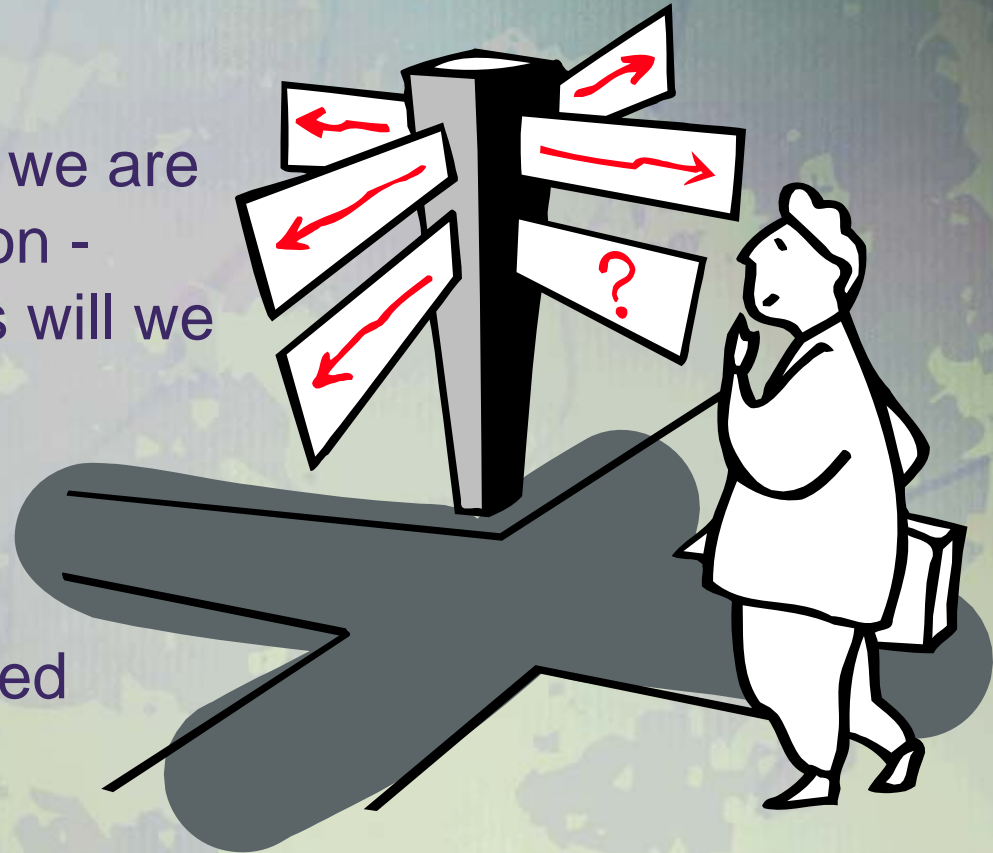
- The term "performance assessment units" used here is meant to describe the development of a linked series of performance assessments based on the three modes of communication, as pioneered by ACTFL. "Performance Assessment Units" used in a pilot project developed by ACTFL through a three-year grant-funded project later evolved into the concept of "Integrated Performance Assessment" (IPA).
- More information about IPA can be found in the manual ***ACTFL Integrated Performance Assessment*** by E. W. Glisan, B. Adair-Hauck, K. Koda, S. P. Sandrock, and E. Swender published in 2003. The manual and IPA workshops are available through **ACTFL**.
- http://www.carla.umn.edu/assessment/vac/CreateUnit/e_1.html

Destination: where are we going?

Roadmap

How will we know we are at our destination - what landmarks will we look for ?

What will we need to get there?



Stages in the Backward Design

Identify desired results-destination/objective

Determine acceptable evidence-assessment/how do we know we arrived

Plan learning experiences and instruction (activities, readings, grammar practice)-what do we need to get to our destination?

Relationship among instruction, assessment and learner outcomes

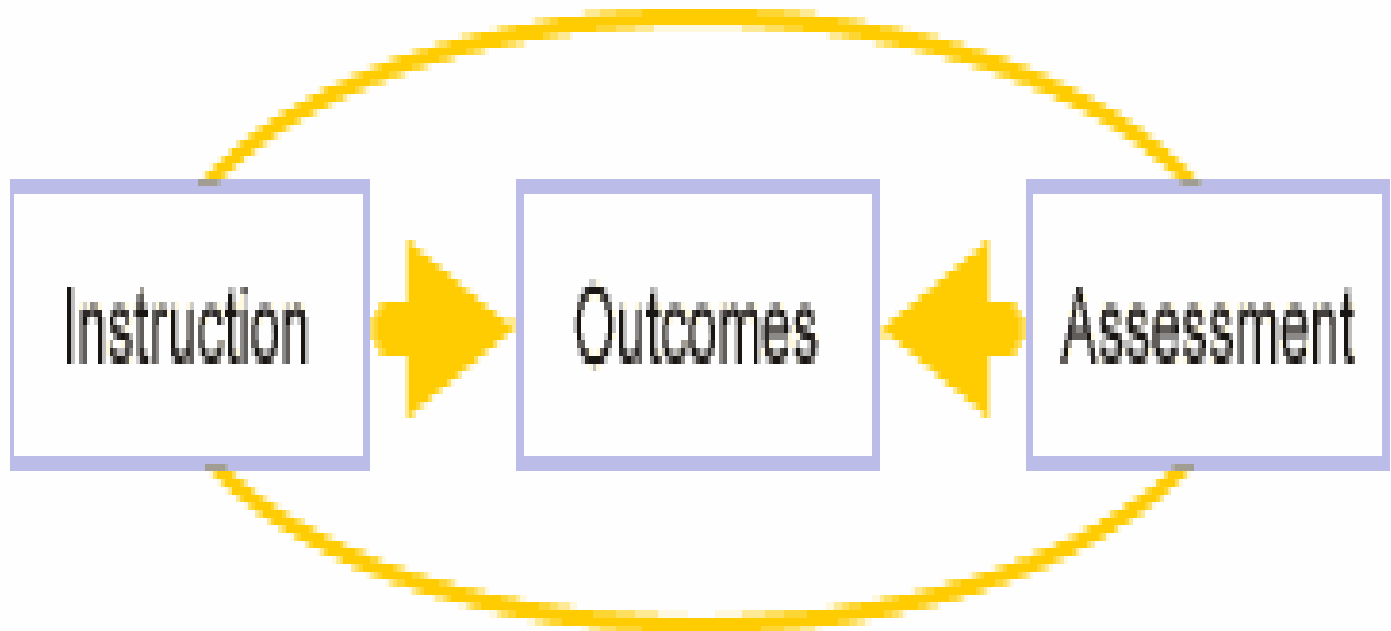
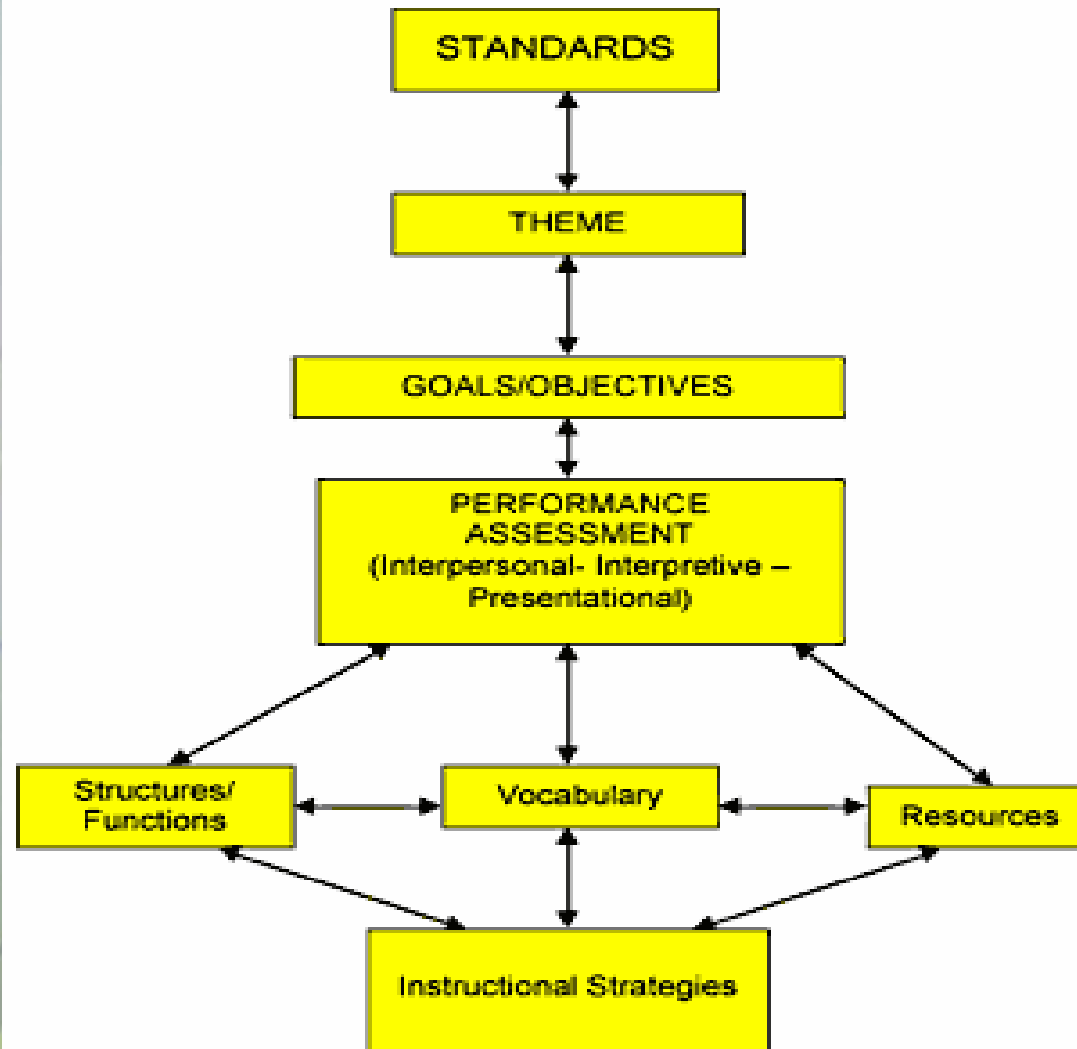


Figure 2 Adapted from Genesee & Upshur (1996, p. 15).



© Donna Clementi, 2003

Identifying desired results

Determine acceptable evidence

- Rating Methods: how to evaluate the evidence
 - objectively - subjectively

Integrated Performance Assessments (IPAs)

- Assessment integrated with curriculum
- Provides real-life context
- Focus is on performance
- Allows for feedback to students

American Council of the Teaching of Foreign Languages. (2003). *Integrated performance assessments*. New York: ACTFL.

- The term "performance assessment units" describes:

a linked series of performance assessments based on the three modes of communication developed by ACTFL that later evolved into the concept of "Integrated Performance Assessment" (IPA).

ACTFL Integrated Performance Assessment by E. W. Glisan, B. Adair-Hauck, K. Koda, S. P. Sandrock, and E. Swender published in 2003

Characteristics of Assessment Units

- Integrated
- Occur over the course of a semester
- Allow for feedback
- Use authentic materials
- Assess all skills under the umbrella of one unifying topic
- Are a culminating assessment

Asking Bigger Questions

a. Clothes _____	I. How do movies reflect culture?
b. Sports _____	II. Can staying fit be fun?
c. Holidays _____	III. How do you define vacation?
d. Careers _____	IV. Does clothing tell a story?
e. Geography _____	V. Who is in charge of my future?
f. Movies _____	VI. Are traditions important

Clothes

- What do your clothes say about you?
- Why do you wear what you wear?
(ethnographic aspect)
- How does your environment affect the clothing you wear?

Sports

- Can we play without winning? (Zero sum game)
- What is a sport?
- How is behavior impacted by the sports we enjoy?
- Why do sports involve balls?
- Is competitiveness in our nature?

Friends

- What makes a good friend good?
- What qualities define a good friendship?
- How does friendship change/start/is different/in different cultures?
- What is the role of a friend?
- How do we prioritize-what is more important, your friends or your family? (in different cultures)

Holidays

- How does the way we celebrate bring our community together?
- How are holidays impacted by the retail industry?

Famous People

- What is fame?
- How long does fame last?

Careers

- What careers are there to choose?
- How does a 2nd language enhance your career choice?
- What is the meaning of career in target culture?
- How important is income to our choice in career?
- Why/how does gender affect your options?
- Will my choices today limit my choices tomorrow?
- What limits your choices?

Friends

- What makes a good friend good?
- What qualities define a good friendship?
- How does friendship change/start/is different/in different cultures?
- What is the role of a friend?
- How do we prioritize-what is more important, your friends or your family? (in different cultures)

UNIT PLAN INVENTORY

Language _____ Level _____ Theme _____

GOALS/OBJECTIVES

- 1.
- 2.
- 3.
- 4.

<p align="center">Communication</p>	<p>Select one: Speaking/writing- (Interpersonal) task</p> <p>Reading/listening (Interpretive) task</p> <p>Writing/Speaking- one way- edited, rehearsed (Presentational) task</p>	<p>Select one: Speaking/writing- (Interpersonal) task</p> <p>Reading/listening (Interpretive) task</p> <p>Writing/Speaking- one way- edited, rehearsed (Presentational) task</p>	<p>Select one: Speaking/writing- (Interpersonal) task</p> <p>Reading/listening (Interpretive) task</p> <p>Writing/Speaking- one way- edited, rehearsed (Presentational) task</p>
<p align="center">Performance Assessment →</p> <p>(Note: The performance assessment tasks are integrated throughout the instructional unit; they are not meant to be given as a whole at the end of the unit.)</p>			
<p align="center">Culture</p>			

Planning for Assessment: What to include over the course of a unit, lesson, or semester?

Includes:

- Multiple measures to provide information: discreet skills tested and reinforced, complementary, alternative
- Purposeful assessments-have face validity: good use of student and instructor time
- Involving test takers and stakeholders in the process
- Integrating teaching, learning and assessment
- Diagnostic: instructional and learning strategies

- On-going (formative) assessments - not a one time event
- Task-based, performance oriented assessments

Questions at the curriculum/unit/lesson/course level

- What do I want my students to know and be able to do?
- What is the evidence I will accept?
- How will I obtain that evidence?

Making an Inventory of assessment practices

Type of test/assessments	How often
Changes and Affirmation	

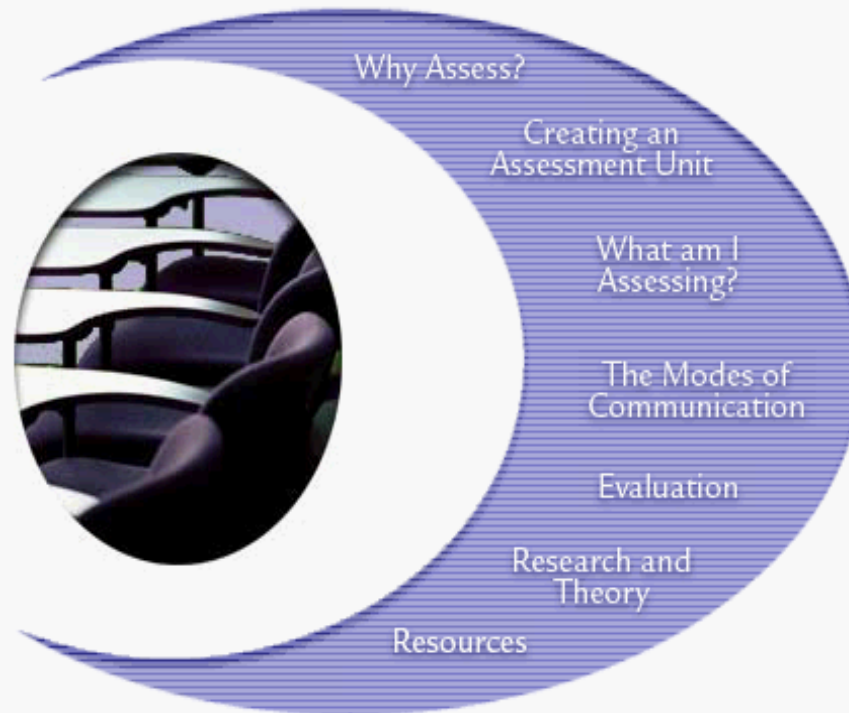
Resources

CARLA Summer Assessment Institute Web Site
<http://www.carla.umn.edu/assessment/si>

CARLA's Virtual Assessment Center (VAC)
<http://www.carla.umn.edu/assessment/vac>

Virtual Item Bank (still in development)
http://dev-carla.umn.edu:16080/assessment/VIB/using_items.html

Virtual Assessment Center



Intermediate Listening (English)

Money, Shopping, Purchases & Numbers

Select Topic...

Listening stimulus:	Intermediate level task requires learners to demonstrate that they can:		
Basic informational, instructional, directional text for a wide public audience	Understand simple questions and answers, simple statements, and simple face-to-face conversations	Understand main gist, main ideas, and some supporting details in authentic material	Detect the mood of the message. Determine to a limited degree the attitude/feelings of speakers or urgency, etc...
<ul style="list-style-type: none"> ▪ Ads (radio & TV commercials) 			
<ul style="list-style-type: none"> ▪ Announcements (airport, train, school PA) 			
<ul style="list-style-type: none"> ▪ Entertainment guides (recorded movie listings, radio entertainment guides) 			
<ul style="list-style-type: none"> ▪ Instructions (cooking shows, home improvement shows, etc) 			
<ul style="list-style-type: none"> ▪ Short Dialogues (staged conversations, movie & TV show clips) 	Short dialogues with questions that focus on understanding simple questions and answers, simple statements, and simple face-to-face interaction. Example: Talking ATM		
<ul style="list-style-type: none"> ▪ Short Narratives (storytelling, academic lectures, how-to television and radio) 			
<ul style="list-style-type: none"> ▪ Technical Reports (tech class lectures, science radio and TV shows) 			
<ul style="list-style-type: none"> ▪ Telephone Communications (conversations, answering machine messages) 		Authentic telephone communications, messages, etc...with questions that focus on gist, main ideas, and supporting details. Example: Cycle Asylum	
<ul style="list-style-type: none"> ▪ Weather forecasts (nightly news, radio) 			

Reading stimulus:	Novice level task requires learners to demonstrate that they can:			
	Recognize most symbols in a phonetic writing system (Cyrillic, hangul, kana, etc.)	Recognize isolated words and expressions	Pick out main ideas and key words in familiar material	Comprehend recombined text using familiar, memorized material
<ul style="list-style-type: none"> ▪ Ads (yellow pages, classifieds) 		Find sample ads in the yellow pages or classified section of the newspaper and ask short-answer questions Examples: Yellow Pages	Find sample ads in the yellow pages or classified section of the newspaper and ask short-answer questions Examples: Yellow Pages	
<ul style="list-style-type: none"> ▪ Entertainment guides (TV, radio, movie guide) 		Use an excerpt of TV/radio listings with basic fill-in-the-blank/short answer questions	Use an excerpt of TV/radio listings with basic fill-in-the-blank/short answer questions	Use an excerpt of TV/radio listings with basic fill-in-the-blank/short answer questions
<ul style="list-style-type: none"> ▪ Figures and captions (comic strips, pictures) 		Use pictures of items and ask matching and identification questions Examples: Holiday Wish List	Use pictures of items and ask matching and identification questions Examples: Holiday Wish List	N/A
<ul style="list-style-type: none"> ▪ Forms (applications, ordering) 		Provide students with an application, registration, or order form and ask them to fill in the information. Find a sample membership application and ask short-answer questions.	Provide students with an application, registration, or order form and ask them to fill in the information. Find a sample membership application and ask short-answer questions.	Provide students with an application, registration, or order form and ask them to fill in the information. Find a sample membership application and ask short-answer questions.
<ul style="list-style-type: none"> ▪ Literary texts (short poems) 				
<ul style="list-style-type: none"> ▪ Manuals (user manuals, instruction booklets) 		Give a short article with tips on car maintenance Example: Winterize Your Car	Give a short article with tips on car maintenance Example: Winterize Your Car	Give a short article with tips on car maintenance Example: Winterize Your Car
<ul style="list-style-type: none"> ▪ Maps (directions to location) 				N/A
<ul style="list-style-type: none"> ▪ Menus (restaurant menu) 				
<ul style="list-style-type: none"> ▪ Schedules/timetables (personal schedule or planner) 		Use a sample personal schedule & ask questions about when that person is free & what activity the person does	Use a sample personal schedule & ask questions about when that person is free & what activity the person does	Use a sample personal schedule & ask questions about when that person is free & what activity the person does
<ul style="list-style-type: none"> ▪ Short narratives (Web or magazine articles) 		Give a short article with healthy eating/lifestyle suggestions and ask comprehension questions Examples: Breakfast Ideas	Give a short article with healthy eating/lifestyle suggestions and ask comprehension questions Examples: Breakfast Ideas	Give a short article with healthy eating/lifestyle suggestions and ask comprehension questions Examples: Breakfast Ideas
<ul style="list-style-type: none"> ▪ Short texts (memos, notes, short letters, invitations) 		Use a short memo that listing various objects and ask comprehension/short-answer questions Examples: Moving Sale	Use a short memo that listing various objects and ask comprehension/short-answer questions Examples: Moving Sale	Use a short memo that listing various objects and ask comprehension/short-answer questions Examples: Moving Sale
<ul style="list-style-type: none"> ▪ Signs 				

Video FAQs Introducing Topics in Language Testing from Glenn Fulcher & Randy Thrasher featured on the International Language Testing Association.

Topics covered by the videos include reliability, validity, test impact, test specifications, writing test items, pre-testing, testing: listening, speaking, reading, writing, statistics, testing for specific purposes.

<http://www.le.ac.uk/education/testing/ilta/faqs/main.html>