Welcome!

"Don’t worry, Howard. The big questions are multiple choice."

Victoria Roberts

ALL RIGHTS RESERVED
http://www.cartoonbank.com

Assessments that Matter, SALRC’    ©The New Yorker October 21, 1991
Frank and Ernest cartoon characters on the last day of school before summer vacation:

Frank: The last day of school year, I always have mixed feelings. I’m excited about starting summer, but it’s also report card day. The problem is, nothing in school is like it is in the outside world. School is mostly true false and multiple choice… but real life is all essay questions.

Bob Thaves: Frank and Ernst cartoon
Assessments that Matter
SALRC Assessment Workshop

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

1) Explore and understand general and classroom assessment principles and models

2) Identify standards, guidelines; align assessment with curricular goals

3) Apply knowledge to develop new /modify current assessment or assessment framework
Today

Big Picture view

Assessment Concepts and Theory

Matching assessment to purpose

Identifying and defining expectations

Guidelines, criteria, standards

Assessment concepts and types – test methods
Assessing for Proficiency

Speaking and Writing (CoWA, CoSA, STAMP, portfolio)

Rating Criteria (Checklists, rubrics and scales)

Work to develop speaking and writing assessment
Tomorrow

Reading and Listening Assessments

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

Integrated Performance Assessments - IPA

Developing a performance assessment unit/topically organized unit

Assessment

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.
Many ways to talk about assessment

1. Purpose
2. Characteristics essential to good tests
3. Type (techniques)
4. Practicality
5. Test taking strategies
1. Purpose and Use

<table>
<thead>
<tr>
<th>Type of test/assessments</th>
<th>Purpose?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why do we give tests?

Purpose(s):

Use(s):
Assessment Purposes

- Grades
- Progress
- Program evaluation
- Instructional
- Research
- Diagnostic
# The Purpose of the Assessment

<table>
<thead>
<tr>
<th>General purpose of the assessment</th>
<th>Specific reason for the assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative</td>
<td>general assessment, placement, exemption, certification, promotion</td>
</tr>
<tr>
<td>Instructional</td>
<td>diagnosis, evidence of progress, feedback to the respondent, evaluation of teaching or curriculum</td>
</tr>
<tr>
<td>Research</td>
<td>evaluation, experimentation, knowledge about language learning &amp; language use</td>
</tr>
</tbody>
</table>

Focus:

Instructional

diagnosis
evidence of progress
feedback to the respondent
evaluation of teaching or curriculum
2. Essential Characteristics

Validity

• Construct – test measures what it is intended to measure
• Content – sample is representative of target domain
• Face – test “looks valid” to test takers
• Washback – close relationship between test & 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 Types

Formative  Summative
Direct  Indirect
Objective  Subjective
Integrative  Discrete-point
Norm-referenced  Criterion-referenced
Achievement  Proficiency
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 in which they are engaged (Black and William, 1998).

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

Purpose is accountability.

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

Indirect

e.g., tour guide
observation

paper
and pencil test
Objective and Subjective

Objective can be scored with reference to a scoring key; does not require expert judgment in scoring

Subjective depends on an informed impression and opinion at time of scoring
Discrete-point and Integrative

Discrete-point – items measure performance over a single set of linguistic features, e.g., multiple choice to test which article or verb is correct.

Integrative – measures knowledge of several features, modes or skills, e.g., dictation to measure listening comprehension, spelling or general proficiency. Using learned vocabulary correctly in a context.

More of a task or procedure than an item.
Norm-referenced and Criterion-referenced

Norm-referenced – compares students with each other, either locally or nationally, e.g., SATs, ACTs, etc. Can establish classroom, regional or state norms; teachers rate on a curve.

Criterion-referenced – determines if students have met certain instructional objectives or criteria.
<table>
<thead>
<tr>
<th>Traditional (Objective)</th>
<th>Authentic (Alternative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice</td>
<td>Interviews</td>
</tr>
<tr>
<td>True false</td>
<td>Essay</td>
</tr>
<tr>
<td>Fill in the blank</td>
<td>Panel discussion</td>
</tr>
<tr>
<td>Cloze</td>
<td>Constructed response</td>
</tr>
</tbody>
</table>
Achievement and Proficiency

Achievement – measures performance/knowledge against a set of instructional goals

Proficiency – measures general knowledge, not linked to specific curricula
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
Academic Prompts

open-ended questions or problems - require students to think critically, not just recall knowledge, and then to prepare a response, product, or performance.

require constructed responses under school or exam conditions.

Often are ill-structured, requiring the development of a strategy. Involve analysis, synthesis, or evaluation.
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
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)."

...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).
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

Purpose of evaluating

Collecting information

Interpreting information

Decision making

Figure 1 Aspects of classroom-based evaluation. Genesee & Upshur (1996, p. 6).
Relationship among instruction, assessment and learner outcomes

Based on Aspects of classroom-based evaluation, Genessee & Upshur (1996, p.15)
Considerations in planning instruction and assessment

- Purpose of Evaluating
  - Collecting Information
  - Needs of Learners
  - Interpreting Information
  - Decision making
Sample Traditional Test: Cloze

• Standard reading comprehension test
• Originally used to assess the relative readability of written materials for school children in the United States.

How it is structured:

• 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.
Assume:

The following cloze test at the end of a unit on nature or environment:
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 has four cubs. One dies soon after birth, and another one is often weak to live long. It is this reason that you usually see two cubs with a lioness. Their looks after them well until they two years old. For the first she brings their food. She eats herself first, then she vomits it 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) _______.

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 usually has four cubs. One dies soon after birth, and another one is often too weak to live long. It is for this reason that you usually see only two cubs with a lioness. Their mother looks after them well until they are two years old. For the first year she brings their food. She eats it herself first, then she vomits it 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.

Discussion

What is your group’s reaction to this test type?
When would it be useful?
Why?
What kind of information does it provide?
What if any issues have you identified?
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 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.

http://www.pgcps.pg.k12.md.us/~croom/types_of_assessment.htm
### Complexity: What do assessments test?

<table>
<thead>
<tr>
<th>Recalled information</th>
<th>Applied Knowledge</th>
</tr>
</thead>
</table>
Performance assessments: key components

1. Tasks - effective in eliciting the performance to be assessed.

2. Rating criteria - evaluate quality of the performance.

   Criteria reflect the relative importance of various aspects of the performance, and are appropriate for the population being assessed.

3. Raters - trained to apply the criteria and can do so consistently.

From CARLA’s Virtual Assessment Center
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;

3. 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.

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.


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).

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.


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

- 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)
Communication Standard

Modes of Communication

Interpersonal mode (spontaneous, negotiated, two-way)
  Speaking, writing (e-mail, text messaging)

Interpretive (one-way, understand and interpret)

Presentational (one-way, audience, formal)
  Speaking, writing
Visual Representation of Anticipated Performance Outcomes As Described in the ACTFL Performance Guidelines for K-12 Learners

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, Vietnamese
Other institutions
First Year Swahili Language

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

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

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.”

## Summary of ACE Credit Recommendations for Official ACTFL OPI Ratings

<table>
<thead>
<tr>
<th>Official OPI Rating</th>
<th>Category I</th>
<th>Category II</th>
<th>Category III</th>
<th>Category IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice High/</td>
<td>*<em>Dutch, English</em>, French, Haitian Creole,</td>
<td>**German, Modern Greek, Hindi, Indonesian</td>
<td>**Cambodian, Czech, Hmong, Hebrew, Hungarian, Lao</td>
<td>**Arabic, Cantonese, Japanese, Korean and Mandarin</td>
</tr>
<tr>
<td>Intermediate Low</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 LD</td>
<td>2 LD</td>
<td>3 LD</td>
<td></td>
</tr>
<tr>
<td>Intermediate Mid</td>
<td>4 LD</td>
<td>5 LD</td>
<td>6 LD</td>
<td></td>
</tr>
<tr>
<td>Intermediate High/</td>
<td>8 LD</td>
<td>8 LD + 2 UD</td>
<td>6 LD + 3 UD</td>
<td></td>
</tr>
<tr>
<td>Advanced Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced Mid</td>
<td>10 LD</td>
<td>8 LD + 4 UD</td>
<td>6 LD + 6 UD</td>
<td></td>
</tr>
<tr>
<td>Advanced High/</td>
<td>10 LD + 2 UD</td>
<td>8 LD + 4 UD</td>
<td>6 LD + 6 UD</td>
<td></td>
</tr>
<tr>
<td>Superior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 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

Using the grids in your packet, decide with your group on benchmarks appropriate for your languages for each of the levels

Year 1

Year 2

Year 3

Year 4
Expected Benchmarks

Year 1  Novice High?
Year 2  Novice High/Intermediate-Low?
Year 3  Intermediate-Low/Mid?
Year 4  Intermediate-Mid/High?
Issues/Discussion

- What is acceptable progress in classes and how should it be measured?

- What common outcomes can be identified at specific levels across languages?
Test Frameworks

Rating Scales and Rubrics
Roadmap

Destination: where are we going?

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
Rating methods

Objective and subjective evaluation
Rating methods

- **Objective – traditional assessments**
  - Pros
  - Cons

- **Subjective – authentic assessments**
  - Pros
  - Cons
Rater reliability

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

Checklists:

• Record observed performance.
• Keep track of progress over time, e.g., inventory of skills at beginning and end of courses.
• Self-assessment
Sample checklist - Receptive oral skills – listening/interpretive

☐ 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
Checklists

• Less informative than scaled rubrics
• Efficient
• Record observed performance
• Simplify rubric construction by specifying non-negotiables*

*(Donna Clementi, 2002)
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 assign one integrated evaluation to a performance, e.g., ACTFL Proficiency Guidelines.

• Analytic rubrics and scales evaluate 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 self)
• 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

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

Types of Rubrics

- Generic
- Combination
- Task-specific
Compare:
Holistic

**ACTFL Performance Guidelines for K-12 Learners**

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

<table>
<thead>
<tr>
<th>Analytic Writing Scale for the Spanish FLIP Program, University of Minnesota*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content (30 pts.)</td>
</tr>
<tr>
<td>Organization (20 pts.)</td>
</tr>
<tr>
<td>Language use/Grammar/ Morphology (25 pts.)</td>
</tr>
<tr>
<td>Vocabulary/Word usage (20 pts.)</td>
</tr>
<tr>
<td>Mechanics (5 pts.)</td>
</tr>
</tbody>
</table>
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*

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Fails to persuade the audience</td>
</tr>
<tr>
<td>1</td>
<td>Attempts to persuade but does not provide sufficient support.</td>
</tr>
<tr>
<td>2</td>
<td>Presents a somewhat persuasive argument but without consistent development and support</td>
</tr>
<tr>
<td>3</td>
<td>Develops a persuasive argument that is well developed and supported.</td>
</tr>
</tbody>
</table>

Creating a Rubric

Step 1: Generate potential dimensions

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

*Donna Clementi
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  
|------------|---------------|
|            | Exceeds  
|            | expectations  
|            | 3 Excellent  
|            | Developing  
|            | Meets  
|            | expectations  
|            | 2 Average    
|            | Beginning  
|            | Not there  
|            | yet          
|            | 1 Needs work |

Assessments that Matter, SALRC’
### Write Benchmark descriptors

| Dimension | Accomplished
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exceeds expectations</td>
<td>Meets expectations</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Excellent</td>
<td>Average</td>
</tr>
</tbody>
</table>

**Assessments that Matter, SALRC’**
Avoid:

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

<table>
<thead>
<tr>
<th>Depth</th>
<th>3 shows depth</th>
<th>2 lacks depth</th>
<th>1 no depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>well-balanced</td>
<td>4 moderately well-balanced</td>
<td>3 not so well-balanced</td>
</tr>
<tr>
<td>2</td>
<td>lacks depth</td>
<td>1 no depth</td>
<td>1 total lack of balance</td>
</tr>
<tr>
<td>1</td>
<td>total lack of balance</td>
<td>1 no depth</td>
<td>1 total lack of balance</td>
</tr>
</tbody>
</table>
Unnecessary negative language:

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creative introduction</td>
<td>Adequate introduction</td>
<td>Boring introduction</td>
<td></td>
</tr>
</tbody>
</table>
American Heritage® Dictionary

http://education.yahoo.com/reference/dictionary/entry/whine

• whine  (hw n, w n)

VERB:
whined, whin·ing, whines

• To utter a plaintive, high-pitched, protracted sound, as in pain, fear, supplication, or complaint.
• To complain or protest in a childish fashion.

NOUN:
• The act of whining.
• A whining sound.
• A complaint uttered in a plaintive tone.
Evaluating Rubrics

Scoring Rubric for Whining

• Language
• Criteria
• Dimensions
• Practicality

An entry from the electronic portfolio of Joseph A. Braun, Jr., Curriculum and Instruction, Coe College
http://coe.ilstu.edu/jabraun/professional/rubrics/whinrub.html
*Source Tillamook School District
Writing Rubrics to Evaluate a Speaking Assessment

According to the number on your sheet, in groups

1. Read the ESL speaking assessment on the handout. Please read the purpose for which you will use this speaking assessment.

2. Make a rubric with criteria that meet the assigned purpose (to consider level, consult your handouts if needed).

4. Listen to the student, and rate the sample using the rubric your group produced.
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.
Appropriateness of the instrument

How appropriate is this speaking assessment for your group’s defined goal?

How appropriate are the tasks?

What about face validity?

Reliability?
## Speaking Assessment Scoring Criteria - Intermediate-Low

<table>
<thead>
<tr>
<th>Task Fulfillment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The speaker demonstrates the ability to:</td>
</tr>
<tr>
<td>Provide responses of sufficient length and substance for the evaluator to apply the scoring criteria.</td>
</tr>
<tr>
<td>Speak about the assigned topics and perform the assigned functions.</td>
</tr>
<tr>
<td>Ask and answer questions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>The speaker:</td>
</tr>
<tr>
<td>Demonstrates an adequately broad vocabulary to address the topics.</td>
</tr>
<tr>
<td>Does not have to resort to excessive repetition of words or phrases.</td>
</tr>
<tr>
<td>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.)</td>
</tr>
</tbody>
</table>
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.

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
Assessing Writing

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

and the bad news

Challenges

• 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
Possible writing assessment tasks


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

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

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](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.

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

How to Use this Manual

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
Sample Portfolio Assessments

http://www.anglit.net/main/portfolio/default.html
Assessing Speaking

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
• Frequently assessed informally
How important is pronunciation?
Berlitz

http://www.visit4info.com/static/advert_pages/30341.cfm?back_page=advertiser_pages/Berlitz
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.

Proficiency Assessments
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/
Llamamos Siglo de Oro a un tiempo en el que España fue la primera potencia mundial. España descubría, exploraba y conquistaba nuevas tierras en el mundo, y en esos períodos florecieron la arte (pintura, escultura y arquitectura) y la literatura (poesía, novela y teatro). Además, los españoles exploraron muchos otros continentes.

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

The item of clothing advertised on this Web page is:

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

Submit
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
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:

Creating a speaking or writing assessment

Using the CoSA or CoWA template, work with a group or alone to:

- develop a writing or speaking assessment for your course
- develop how you will rate the assessment.
<table>
<thead>
<tr>
<th>I. CoSA Theme</th>
<th>Segments</th>
<th>II. Topic</th>
<th>III. Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provide students with rich description of the general setting in which they are asked to imagine themselves speaking in L2.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>What is the general setting? What motivates the speaker to speak L2?</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Warm-up</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Is the speaker in a plausible situation where he/she can respond with personal experiences?</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes/no</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>With how many characters does the speaker interact?</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Who are the characters (age, status, occupation)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Segment 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Segment 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Segment 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Segment 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Segment 5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wind-down</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Minnesota Language Proficiency Assessments**

**Contextualized Speaking Assessment (CoSA) Development Worksheet: Level: Intermediate (Part 1)**
<table>
<thead>
<tr>
<th>Segments</th>
<th>IV. Functions</th>
<th>V. Task</th>
<th>VI. Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asking questions</td>
<td>Specify the speaker's task</td>
<td>Write the prompt in the target language</td>
</tr>
<tr>
<td></td>
<td>Making suggestions</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Describing in areas that pertain to self and immediate environment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Narrating based on a sequence of pictures</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Express likes and dislikes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Warm-up</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Segment 1</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Segment 2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Segment 3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Segment 4</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Segment 5</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Wind-down</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
Assignment

• Identify a text, podcast or other reading or listening related to a topic you teach in your course for use in developing an assessment.

• Ideas: Fables, newspaper articles relating to the topic, segments of movies used in class
Yesterday

Purpose and Use of assessments

Assessment terminology and characteristics

Guidelines and level expectations

Rubrics- develop and apply

Proficiency – writing and speaking
Today
Assessing Reading and Listening
Integrated Performance Assessments
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

• Should be tested when speaking cannot be

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.

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

<table>
<thead>
<tr>
<th>Context/Theme</th>
<th>Text Type(s)</th>
<th>Topic(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Returning from a trip</td>
<td>Informal conversation among peers; PA announcement</td>
<td>Travel, transportation, airport information, vacation plans, invitations, instructions</td>
</tr>
<tr>
<td>Café; restaurant</td>
<td>Informal and formal conversations</td>
<td>Travel, leisure, work, friends, money, ordering food and beverages, likes and dislikes</td>
</tr>
<tr>
<td>Making plans for the day</td>
<td>Informal conversations</td>
<td>Leisure, transportation, food, film, invitations, preferences</td>
</tr>
<tr>
<td>Doing errands</td>
<td>Informal and formal conversations, radio broadcast</td>
<td>Directions, news event, music, housing, post office transactions, money, appointments, instructions</td>
</tr>
<tr>
<td>Friends and family</td>
<td>Informal conversation</td>
<td>Greetings, introductions, studies, leisure, family, health, sports</td>
</tr>
<tr>
<td>Shopping</td>
<td>Informal conversations, news report, announcement, commercial</td>
<td>Money, clothing, colors, product description, leisure, preferences, current events</td>
</tr>
</tbody>
</table>
Listen to the sample from
http://dev-carla.umn.edu/assessment/VIB/

In what other ways could you assess the segment?
Assessing Reading

Reading Comprehension
Three-level process 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

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

## Reading Activities/Assessments

<table>
<thead>
<tr>
<th>Organize</th>
<th>Demonstrate</th>
<th>Process</th>
<th>Create</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job application</td>
<td>Ad</td>
<td>Discussion</td>
<td>Letter New ending</td>
</tr>
<tr>
<td>Map</td>
<td>Critique</td>
<td>Journal</td>
<td>Updated version</td>
</tr>
<tr>
<td>Outline</td>
<td></td>
<td>Retelling</td>
<td>Video</td>
</tr>
<tr>
<td>Timeline</td>
<td></td>
<td>Script</td>
<td></td>
</tr>
<tr>
<td>Travel log</td>
<td></td>
<td>Socratic seminar</td>
<td></td>
</tr>
<tr>
<td>Venn diagram</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Integrated Performance Assessments (IPAs)

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

• 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).

Integrated performance assessments (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](http://www.carla.umn.edu/assessment/vac/CreateUnit/e_1.html)
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.”

Roadmap

Destination: where are we going?

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 the destination?
Identifying desired results

Determine acceptable evidence: What

Rating Methods: How to best evaluate the evidence

objectively - subjectively

Determine what is needed to arrive at the result: instruction, practice, resources,
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
Sample IPA: Can One Person Make a Difference?

Level: IH-AL

Interpretive Task:

Divide the class into groups of five or six students. Assign each group the name of a person or organization from the target culture that works to make the world a better place to live. For example, French students might investigate Medecins Sans Frontieres (Doctors Without Borders). Students then research their person or organization on the Internet in order to find out the following information:
• http://www.doctorswithoutborders.org/
• http://www.ifrc.org/
• http://www.salvationarmy.org
• http://www.savethechildren.org/
• http://www.unicef.org/
• How was this group founded and when and by whom?
• What are the goals for this group?
• Give examples of the work/successes of the group.
• What are the current projects of this group?
• Interpersonal Task:

PART ONE: After students have gathered information about their organization, they will discuss with their group the information that they found in order to determine which details are the most important to share with others. When they have determined what information to share, they will move into a jigsaw activity sharing their information with others who researched different organizations.
Presentational Task:

Based on the information presented in the group discussions (interpersonal task), students are to write an essay stating which organization they would support either financially or through personal involvement and why. They may also choose to write why they would not support any of the organizations financially or through personal involvement. The essay should include background information on the organization and reasons why the organization makes (or doesn’t make) a difference in the world.

## Asking Bigger Questions

<table>
<thead>
<tr>
<th>a. Clothes</th>
<th>I. How do movies reflect culture?</th>
</tr>
</thead>
<tbody>
<tr>
<td>b. Sports</td>
<td>II. Can staying fit be fun?</td>
</tr>
<tr>
<td>c. Holidays</td>
<td>III. How do you define vacation?</td>
</tr>
<tr>
<td>d. Careers</td>
<td>IV. Does clothing tell a story?</td>
</tr>
<tr>
<td>e. Geography</td>
<td>V. Who is in charge of my future?</td>
</tr>
<tr>
<td>f. Movies</td>
<td>VI. Are traditions important</td>
</tr>
</tbody>
</table>
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?
PART TWO:

Jigsaw activity: Form new groups of students so that the new groups are composed of representatives from each of the original research groups. In each new group, the students will share the information they learned about their organization. Other students in the group are encouraged to ask questions to gain more information about each organization.
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?
• Would you want to be famous?
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?
**UNIT PLAN INVENTORY**

Language ___________________ Level ___________________ Theme ___________________

**GOALS/OBJECTIVES**

1. 
2. 
3. 
4. 

<table>
<thead>
<tr>
<th>Communication</th>
<th>Select one: Speaking/writing- (Interpersonal) task</th>
<th>Select one: Speaking/writing- (Interpersonal) task</th>
<th>Select one: Speaking/writing- (Interpersonal) task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reading/listening (Interpretive) task</td>
<td>Reading/listening (Interpretive) task</td>
<td>Reading/listening (Interpretive) task</td>
</tr>
<tr>
<td></td>
<td>Writing/Speaking- one way- edited, rehearsed (Presentational) task</td>
<td>Writing/Speaking- one way- edited, rehearsed (Presentational) task</td>
<td>Writing/Speaking- one way- edited, rehearsed (Presentational) task</td>
</tr>
</tbody>
</table>

**Performance Assessment**

(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.)

| Culture       |                                                |                                                |                                                |

**Used with permission, Donna Clementi and Helena Curtain, 2003**

*Addition by MDE World Language QTN, June 2004*
<table>
<thead>
<tr>
<th>Reflection:</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structures</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
</tr>
<tr>
<td></td>
<td>Materials/Resources</td>
</tr>
<tr>
<td>Major Lesson</td>
<td>Beginning, Middle, End</td>
</tr>
</tbody>
</table>

Used with permission, Donna Clements and Helena Curtain, 2003
Planning for Assessment
What should be included 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 strategy
• On-going (formative) assessments not a one time event
• Task-based, performance oriented assessments
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)
## Intermediate Listening (English)

### Money, Shopping, Purchases & Numbers

<table>
<thead>
<tr>
<th>Listening stimulus:</th>
<th>Intermediate level task requires learners to demonstrate that they can:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic informational, instructional, directional text for a wide public audience</td>
<td>Understand simple questions and answers, simple statements, and simple face-to-face conversations</td>
</tr>
<tr>
<td>- Ads (radio &amp; TV commercials)</td>
<td></td>
</tr>
<tr>
<td>- Announcements (airport, train, school PA)</td>
<td></td>
</tr>
<tr>
<td>- Entertainment guides (recorded movie listings, radio entertainment guides)</td>
<td></td>
</tr>
<tr>
<td>- Instructions (cooking shows, home improvement shows, etc)</td>
<td></td>
</tr>
<tr>
<td>- Short Dialogues (staged conversations, movie &amp; TV show clips)</td>
<td>Short dialogues with questions that focus on understanding simple questions and answers, simple statements, and simple face-to-face interaction. Example: Talking ATM</td>
</tr>
<tr>
<td>- Short Narratives (storytelling, academic lectures, how-to television and radio)</td>
<td></td>
</tr>
<tr>
<td>- Technical Reports (tech class lectures, science radio and TV shows)</td>
<td></td>
</tr>
<tr>
<td>- Telephone Communications (conversations, answering machine messages)</td>
<td>Authentic telephone communications, messages, etc., with questions that focus on gist, main ideas, and supporting details. Example: Cycle Asylum</td>
</tr>
<tr>
<td>- Weather forecasts (nightly news, radio)</td>
<td></td>
</tr>
</tbody>
</table>
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/ilda/faqs/main.html
Thank you!