

***MEETING THE DIVERSE CHALLENGES OF
ASSESSMENT IN
COLLEGE FOREIGN LANGUAGE PROGRAMS***

John M. Norris

University of Hawai'i at Mānoa

Association of Departments of Foreign Languages
Summer Seminar West

Albuquerque, New Mexico
June 11, 2004

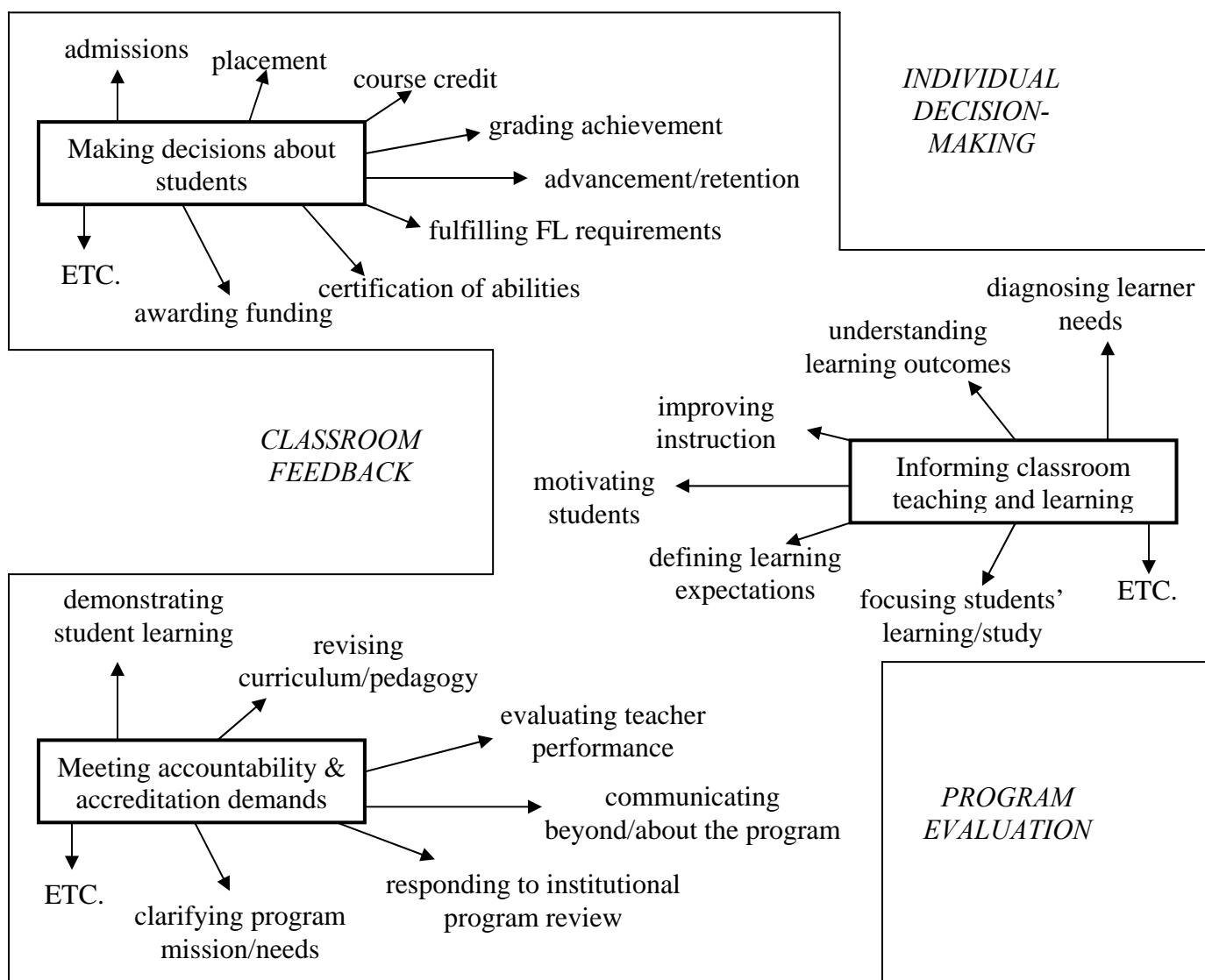
1. Introduction: The challenge of assessment in college FL education

What is educational assessment?

Popham (2000): “A process by which educators use students’ responses to specially created or naturally occurring stimuli in order to make inferences about students’ knowledge, skills, or affective status” (p. 3).

Norris: The purposeful use of tests and related instruments and procedures for gathering trustworthy information about learners *and making decisions and taking actions on the basis of that information.*

What are the impetuses for assessment in college FL education? Why do we assess?



What are the principal challenges that college FL programs face in responding to these impetuses for assessment?

- Traditions of FL assessment have focused on the ‘HOW’ of measuring language rather than the ‘WHY’ of educational assessment within FL programs (e.g., Barnwell, 1996; Spolsky, 1995)
- Professional responsibility for, and development of capacity in, assessment has often lacked within college FL programs (e.g., Grosse, 1993; Omaggio, 1986, 2001).
- Assessment R&D (e.g., within the language testing discipline) has only minimally addressed the many impetuses and needs of assessment within college FL programs.

Spolsky (2000) on history of assessment articles in the *MLJ*: any “serious contributions to the field of language testing” (p. 536)?

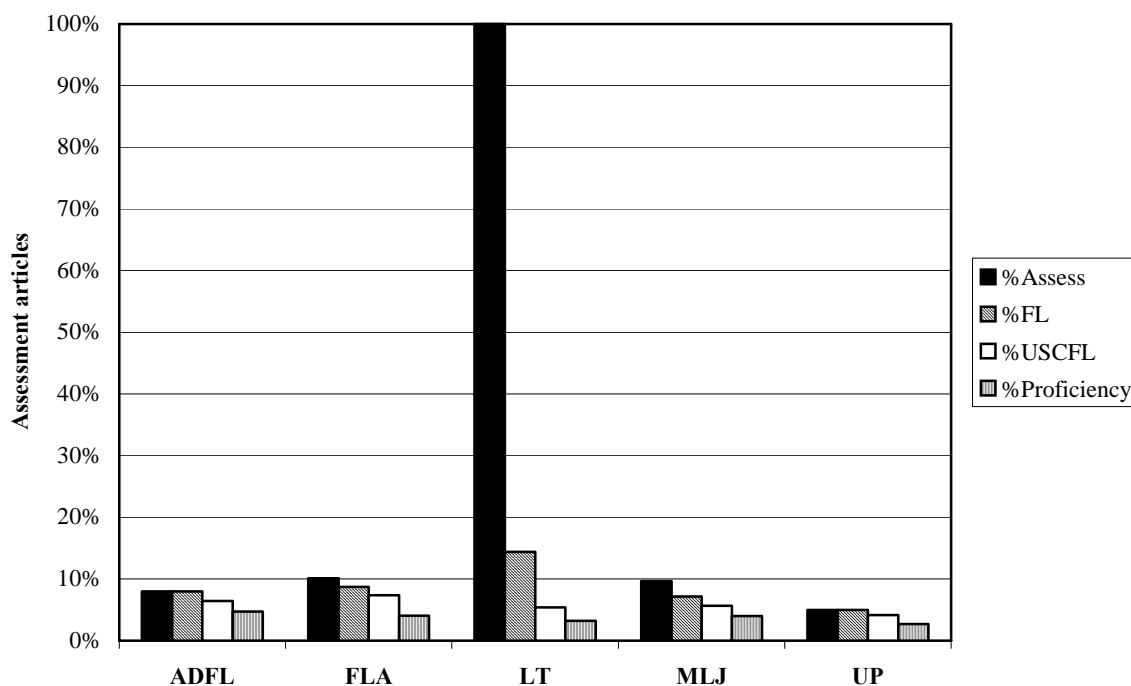


Figure 1. Percentage and type of FL assessment articles in five journals, 1984-2002

Note: Abbreviations are as follows: ADFL = *ADFL Bulletin*; FLA = *Foreign Language Annals*; LT = *Language Testing*; MLJ = *Modern Language Journal*; UP = *Die Unterrichtspraxis*; %Assess = percentage of assessment-focused articles; %FL = percentage of total non-English FL assessment articles; %USCFL = percentage of assessment articles related to the U.S. college foreign language context; %Proficiency = percentage of articles related to assessment based on the ACTFL Proficiency Guidelines.

(Source: Norris, 2004, p. 64)

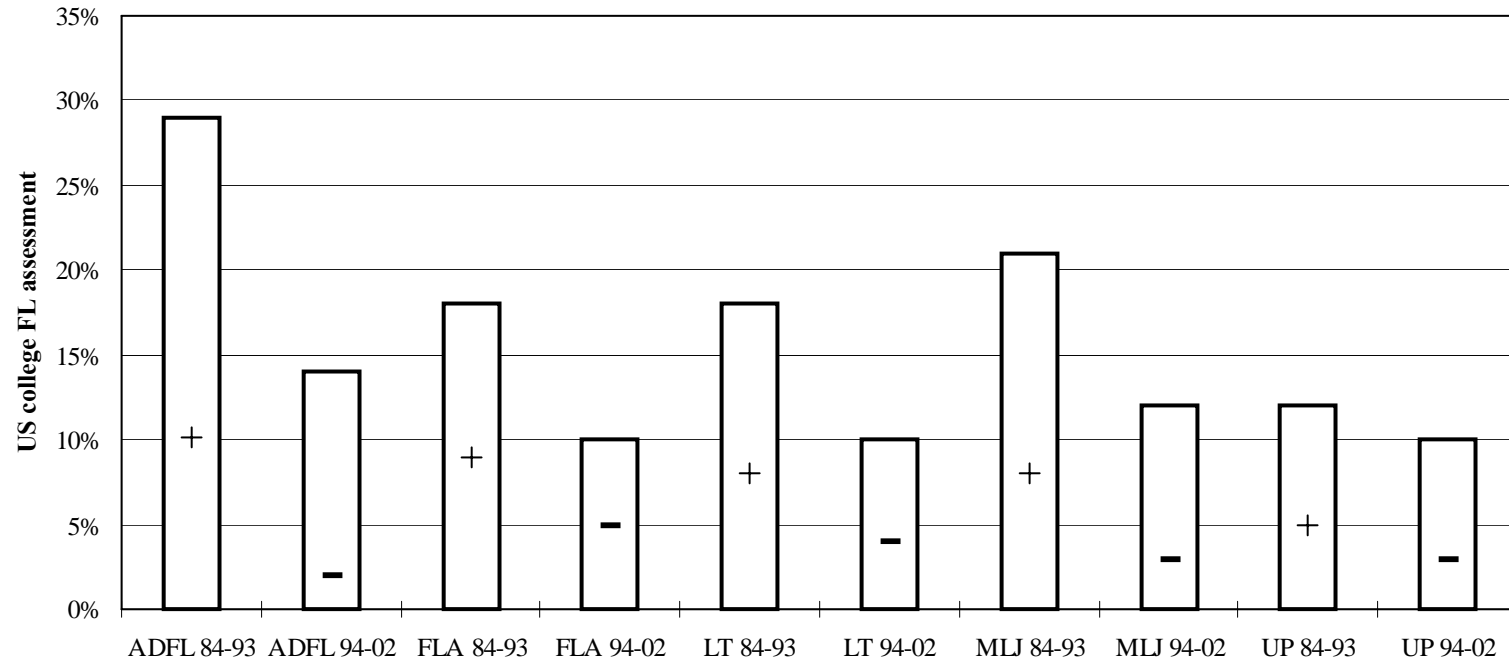


Figure 2. High, low, and average yearly percentage of articles on U.S. college FL assessment, 1984-1993 and 1994-2002

Note: Abbreviations are as follows. ADFL = *ADFL Bulletin*; FLA = *Foreign Language Annals*; LT = *Language Testing*; MLJ = *Modern Language Journal*; UP = *Die Unterrichtspraxis*.

(Source: Norris, 2004, p. 67)

As a result, assessment in college FL programs may suffer from...

- Lack of alignment with (or support for) program-specific curriculum & instruction
- Uncertain or absent purposes
- Negative washback on teaching & learning
- Overgeneralization of the meaning of assessment scores and outcomes
- Inappropriate and even deleterious uses of assessments
- and other problems...

An example: Using ACTFL Guidelines oral proficiency ratings in a German program

Table 1. ACTFL Guidelines oral proficiency ratings for students in the GUGD

	<u>GU German curricular level completed</u>						
<u>ACTFL level</u>	1.1 (3 hrs.) N=7	1.2 (6 hrs.) N=26	2.1 (9 hrs.) N=19	2.2 (12 hrs.) N=39	3.1 (15 hrs.) N=20	3.2 (18 hrs.) N=12	4.1 (22 hrs.) N=5
Novice-High	1			1			
Intermediate-Low	4	7	5	10	2		
Intermediate-Mid	2	15	10	18	4		
Intermediate-High		4	4	7	10	7	
Advanced				2	3	3	
Advanced-High ^a				1	1	2	4
Superior							1
Median rating	IL	IM	IM	IM	IH	IH	AH
Mode rating	IL	IM	IM	IM	IH	IH	AH

^aBecause a number of ratings were completed before the publication of the revised ACTFL (1999) Proficiency Guidelines for Speaking, where the Advanced-Low and Advanced-Mid designations were added to the proficiency scale, all ratings are reported here according to the ACTFL (1986) Proficiency Guidelines.

(Source: Norris & Pfeiffer, 2004, p. 577)

Summary challenge: How can we clarify the ‘why’ of college FL assessment, develop/adopt/adapt instruments and procedures that can be articulated with assessment purposes, and ensure the quality of our assessments in fulfilling specific roles within our specific FL educational contexts? How can we treat assessment as an essential component of college FL education, along with curriculum and instruction?

➤ Key References

- Barnwell, D. P. (1996). *A history of foreign language testing in the United States*. Tempe, AZ: Bilingual Press.
- Grosse, C. U. (1993). The foreign language methods course. *Modern Language Journal*, 77, 303-312.
- Norris, J. M. (2004). *Validity evaluation in foreign language assessment*. Unpublished doctoral dissertation. Honolulu, HI: University of Hawaii at Manoa.
- Norris, J. M., & Pfeiffer, Peter. (2004). Exploring the uses and usefulness of ACTFL Guidelines oral proficiency ratings and standards in college foreign language departments. *Foreign Language Annals*, 37(1), 572-581.
- Omaggio, A. C. (1986). *Teaching language in context: Proficiency-oriented instruction*. Boston: Heinle & Heinle.
- Omaggio-Hadley, A. C. (2001). *Teaching language in context* (3rd ed.). Boston: Heinle & Heinle.
- Spolsky, B. (2000). Language testing in *the Modern Language Journal*. *Modern Language Journal*, 84(4), 536-552.
- Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.

2. Meeting the challenge: Specification of intended uses for assessment

A. Rationale: Before developing instruments and procedures for assessment (i.e., the ‘how’ of assessment), the programmatic role(s) of each assessment should be specified in terms of what it is intended to accomplish. With a clear specification of intended assessment use, we are provided both with a mandate for development and a foundation for evaluation of assessment practices. Without a clear specification of intended assessment use, we risk *useless* assessment practices.

Bachman & Palmer (1996): “The most important consideration in designing and developing a language test is the use for which it is intended” (p. 17).

B. Process: Components and steps in specifying intended uses for assessment

1. Clarify intended uses for all existing and needed assessments

- Who are the test users—teachers, students, administrators, institutions, reviewers, etc.?
- What information do they need—student knowledge, ability, affect, progress, outcomes, proficiency, etc.?
- What is the purpose of the test—how will the information actually be used?
- What are the intended consequences of the test—what do we hope happens as a result of the use of the information?

Methods: observation, surveys, discussion, values clarification

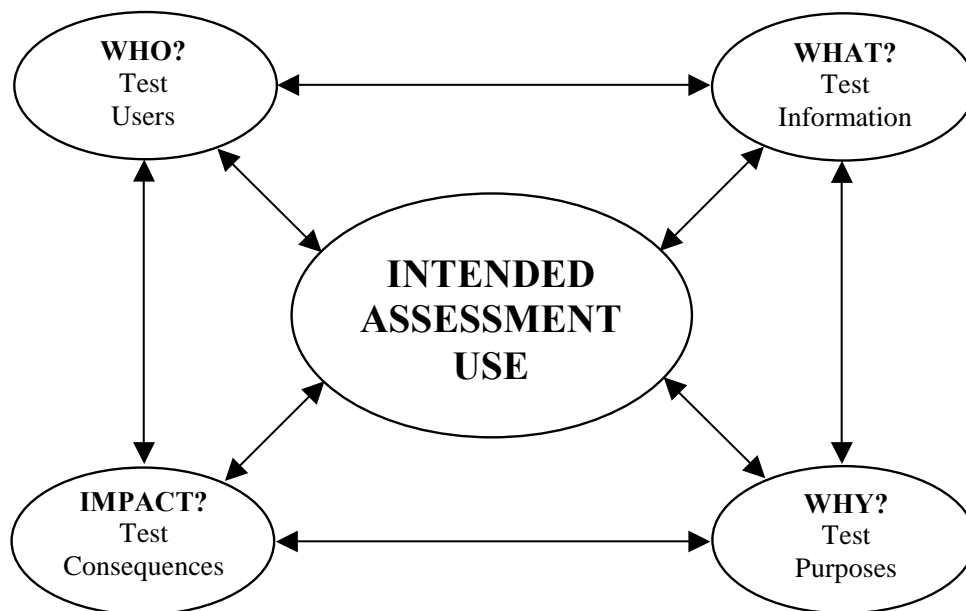


Figure 3. Specifying intended uses for assessment

2. Identify constraints in the assessment context

- time allotted for test administration, scoring, and score reporting
- logistics of test administration, scoring, and score reporting
- resources/materials available
- teacher preparation/training for developing the test
- teacher preparation/training for administering the test
- teacher preparation/training for scoring the test
- public acceptance/communication
- legal issues (high stakes test?)
- fairness/equivalency concerns
- security
- ETC.

3. Seek consensus on final assessment use specifications

- Key program constituents negotiate final specifications
- Prioritize assessment development needs
- Dispense with 'useless' assessments

C. Example: Assessment in the Georgetown University German Department (GUGD)

Developing Multiple Literacies undergraduate German language program (circa 1999):

<p><u>Curriculum</u></p> <ul style="list-style-type: none"> • GU learner needs-based • Integrated Content + L2, all years • Advanced literacy orientation 	<p><u>Instruction</u></p> <ul style="list-style-type: none"> • Genre, task, text driven • In-house materials & pedagogy • All faculty teach all levels 	<p><u>Assessment?</u></p> <ul style="list-style-type: none"> • Purposeful? • Curriculum-based? • Learning-relevant? • Used, beyond grading? • ANY GOOD?
---	--	---

Outcomes from specification of intended assessment uses in the GUGD

“Taken together, these documents are intended to guide not only the development and implementation, but also the evaluation and revision of all quizzes, tests, examinations, written and oral performances, and other forms of assessment which play an integral role in the success of the GUGD’s educational efforts.” (Assessment policies in the GUGD, rev. August, 2002)

1. Departmental assessment policies

Summary of assessment policies in the Georgetown University German Department

1. All assessment focuses on students’ abilities to use the language for communication.
2. Students should have ample opportunity to develop and demonstrate the content, linguistic knowledge, and performance abilities reflected in classroom and cross-curricular assessments.
3. Assessment tasks are gauged at the particular level of instruction (as determined by the curriculum) and reflect suitable expectations in terms of L2 acquisition.
4. Assessments require clear scoring criteria that are communicated to and agreed upon by students and teachers.
5. Assessments provide a clear rationale for the weighting of content and language performance elements in scoring, grading, and feedback.
6. Students are made aware of the specific uses for assessment outcomes.
7. The overall balanced development of accuracy, complexity, and fluency of L2 performance is emphasized within classroom assessment.
8. Multiple assessment formats are required for addressing holistic as well as specific aspects of student learning.
9. Assessment outcomes are conveyed to students and other score users in the form of rich and interpretable feedback in addition to simple grades or scores.
10. Feedback and scores enable both criterion-referenced interpretations, in terms of instructional expectations, and individual-referenced interpretations, in terms of individual student needs and progress.
11. Assessment and pedagogy emphasize the curriculum in similar ways, and one supports the other.
12. Assessment reflects a balanced emphasis on communication modes (listening, speaking, reading, writing), as appropriate to a given course at a given curricular level.
13. Target performance tasks serve as an organizing principle for both assessment and pedagogy.
14. Teachers require training in the use and interpretation of assessment practices appropriate to curriculum and instruction.
15. Teachers and administrators cooperate in developing, using, and revising assessments, and coordination of assessment is designated to particular individuals at each level of use (classroom, curricular-level, and cross-curricular).

2. Departmental assessment development/use priorities (circa 1999)

<i>Assessment</i>	<i>Intended Use</i>
New placement exam	Curriculum-based learner placement
Task-based writing assessment	Understanding student learning outcomes, improving curriculum & instruction
German FL proficiency test	External accountability, student uses
Measures of accuracy & fluency development	Researching instructional effectiveness

Table 1. Examples of GUGD assessment uses at distinct program levels

Specification level	Uses	Assessments
Cross-curricular	--placement	--curriculum-based placement exam
	--curriculum & instruction formative/summative evaluation	--end-of-semester oral & written performance assessment
	--curriculum-external communication, accountability	--proficiency exam (<i>German Speaking Test</i>)
Level-specific (i.e., determined for each curricular semester and year)	--institutional grading requirement	--task-based writing (e.g., letters, briefs, narratives, precis, etc.) --task-based speaking (e.g., presentations, interviews, debates, etc.) --mid-term and final exams --construct measures of accuracy & fluency acquisition
	--on-going curriculum & instruction improvement	
	--syllabus-level performance feedback to teachers and students	
	--curricular alignment across common courses/sections	
	--research on instructional effectiveness	
Classroom-based	--short-term diagnosis and student motivation	--daily quizzes --conferences --journals --homework --self-assessment --class participation rubrics
	--focused feedback on discrete learning elements	

3. Summary specifications of intended uses for assessment (short examples)

GUGD Placement exam: Intended use

Who: Faculty/administrative decision makers; also students

What: Estimate of incoming students' curriculum-related German L2 knowledge/abilities

Why: Placement into curricular level acknowledging abilities and addressing learning needs

Impact: Efficient/effective teaching and learning for learners grouped by similar ability and need

Constraints

- Wide range of learner and curricular ability levels
- Administration time (<2 hr.)
- Scoring time (same day)
- Decision-making efficiency
- Transportability (off-site administration), migration to computer-based administration

GUGD Task-based writing assessment: Intended use

Who: Faculty, instructors, curriculum developers

What: Representative sample of writing performance abilities at the end of each curricular level

Why: Understanding student achievement of targeted abilities for improvement of C&I

Impact: Enable feasible curricular expectations supported by effective pedagogy

Constraints

- Explicitness of curricular expectations
- Consensus on 'prototypical' performance tasks by level
- Competing uses for, demands on assessment
- End-of-semester timing
- Learner investment, understanding of task expectations

➤ **Key References**

Bachman, L., & Palmer, A. (1996). Language testing in practice. Oxford: Oxford University Press.

Norris, J. M. (2000). Purposeful language assessment. English Teaching Forum, 38(1), 18-23.

Norris, J. M. (2002). Specification of intended test use. Flagstaff, AZ: Unpublished technical report for the *National Flagship Language Initiative*, Diagnostic Assessment Procedure.

➤ **Additional Resources**

Georgetown University German Department, *Developing multiple literacies* web site:

<http://www.georgetown.edu/departments/german/programs/curriculum/>.

3. Meeting the challenge: Development and implementation of assessments

A. Rationale: In order for assessments to meet intended uses within college FL programs, the ‘how’ (methods, instruments, procedures) must be articulated with the ‘why’ (users, information, decisions, actions, consequences) of assessment. The specification of intended uses for assessment provides a basis for selecting or developing, and then operationalizing, appropriate tests and related practices.

Norris (2004): “How assessment gets done right—what qualities constitute appropriate, useful, effective assessments—will depend, then, on a thorough understanding of why assessment is being done and what it is intended to accomplish, and practices will have to be designed accordingly” (p. 33).

B. Process: Developing, adopting, adapting, and implementing assessments

1. Consider the qualities required of assessments within the specification of intended use

- focus on the type, depth, and amount of information needed from instruments and procedures
- set limits on feasibility of instruments and procedures, based on absolute constraints of your program context
- avoid creating roles for assessment where they do not exist; stick to the intended uses

Table 2. Example desired qualities of language tests for NFLI learner selection assessment

<u>Language tests used for NFLI learner selection should be:</u>	<u>Evaluation:</u>
a. <u>interpretable</u> as a trustworthy indication of proficiency in speaking/listening/reading in the target language	
b. <u>available</u> in comparable formats for all of the NFLI target languages	
c. <u>appropriate</u> for the variety of potential program applicants (heritage, expatriate, and classroom learners)	
d. <u>administrable</u> on-site at Flagship host institutions or wherever potential students may apply from	
e. <u>scorable</u> , both efficiently and consistently	
f. <u>relevant</u> to the Flagship language learning curricula and objectives	
g. <u>recognizable</u> by Flagship practitioners and other interested constituents (such as federal funding agencies)	
h. <u>affordable</u> within the budgetary constraints of the individual Flagship programs	

2. Review existing options for instruments and procedures

- survey the assessment literature and related resources
- consider assessments used within other programs
- solicit input from local and external experts

Table 3. Examples of literature related to recent language and higher education foci

Assessment focus	Example References
Accountability assessment	Shavelson, R., & Huang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. <i>Change</i> , 35(1), 10-19.
CBTs, CATs, & web-based language testing	Special issue. (2001). Computer-assisted language testing. <i>Language Learning & Technology</i> , 5(2).
Classroom assessment	Angelo, T. A., & Cross, P. K. (2001). <i>Classroom assessment techniques: A handbook for college teachers</i> (2nd ed.). San Francisco: Jossey-Bass. Brookhart, S. M. (1999). The art and science of classroom assessment: The missing part of pedagogy. <i>ASHE-ERIC Higher Education Report</i> , 27(1). Eric Document ED432937. (Available online through Cline Library EBSCO Host subscription .)
Criterion-referenced language testing	Brown, J. D., & Hudson, T. (2002). <i>Criterion-referenced language testing</i> . New York: Cambridge University Press. Davidson, F., & Lynch, B. K. (2002). <i>Testcraft: A teacher's guide to writing and using language test specifications</i> . New Haven: Yale University Press.
Measurement in second language acquisition research	Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty, & M. H. Long, (Eds.), <i>Handbook of second language acquisition</i> (pp. 717-761). London: Blackwell. Bachman, L., & Cohen, A. (1998). <i>Interfaces between second language acquisition and language testing research</i> . Cambridge: Cambridge University Press.
Performance assessment	McNamara, T. (1996). <i>Measuring second language performance</i> . New York: Longman. Norris, J. M., Brown, J. D., Hudson, T. D., & Yoshioka, J. K. (1998). <i>Designing second language performance assessments (Technical Report #18)</i> . Honolulu, HI: University of Hawai'i Press.
Portfolio assessment	Barnhardt, S., Kevorkian, J. A., & Delett, J. S. (1998). <i>Portfolio assessment in the foreign language classroom</i> . Washington, DC: National Capital Language Resource Center. Delett, J. S., Barnhardt, S., & Kevorkian, J. A. (2001). A framework for portfolio assessment in the foreign language classroom. <i>Foreign Language Annals</i> , 34(6), 559-568.
Standards setting in assessment	Cizek, G. J. (Ed.). (2001). <i>Setting performance standards: Concepts, methods, and perspectives</i> . Mahwah, NJ: Lawrence Erlbaum Associates.
Student learning outcomes assessment	Liskin-Gasparro, J. E. (1995). Practical approaches to outcomes assessment: The undergraduate major in foreign languages and literatures. <i>ADFL Bulletin</i> , 26(2), 21-27. Maki, P. (2002). Using multiple assessment methods to explore student learning and development inside and outside of the classroom. <i>NetResults</i> (January), Retrieved January 8, 2003, from http://www.naspa.org/NetResults/article.cfm?ID=2558 .
Task-based language assessment	Norris, J. M. (Ed.) (2002). Special issue: Task-based language assessment. <i>Language Testing</i> , 19(4).

A few useful web sources for assessment ideas (see also resources in section 5):

electronic assessment tools (reviews, links, free online assessment tools, etc.):

<http://www4.nau.edu/assessment/main/electronic/electronic.htm#tools>

<http://www.wcer.wisc.edu/salgains/instructor/default.asp>

learner-centered assessment (examples, handouts, etc.):

<http://www4.nau.edu/assessment/oaainfo/workshop/workshop.htm#learner>

language and other test data-bases:

<http://www.cal.org/topics/tests.html>

<http://www.ets.org/>

<http://ericae.net/testcol.htm>

3. Select assessment instruments and procedures that best fit desired/required qualities

- keep resource constraints in mind
- adopt existing assessments where they best meet intended uses, not simply for ease of use; all adoption will require some degree of adaptation (to your learners, your program needs, your constraints)
- adapt carefully, as changes in instruments and procedures may denigrate assessment qualities
- develop instruments and procedures where existing assessments do not meet desired qualities (seek assistance)

4. Operationalize assessment instruments and procedures (see Brown, 2004; Brown & Hudson, 2002; Davidson & Lynch, 2002)

- test & item writing
- test formatting
- administration procedures & guidelines
- scoring procedures & guidelines
- decision-making, feedback, and other use procedures & guidelines
- communication about assessment with stakeholders (e.g., students) before, during, and after assessment

5. Pilot-test and revise

- try out the assessment with representative students, administrators, scorers
- consider trials with first-language users as well as learners, for critical feedback
- track expected/unexpected outcomes and revise as needed
- pilot all of the features in #4 above (not just the test items)

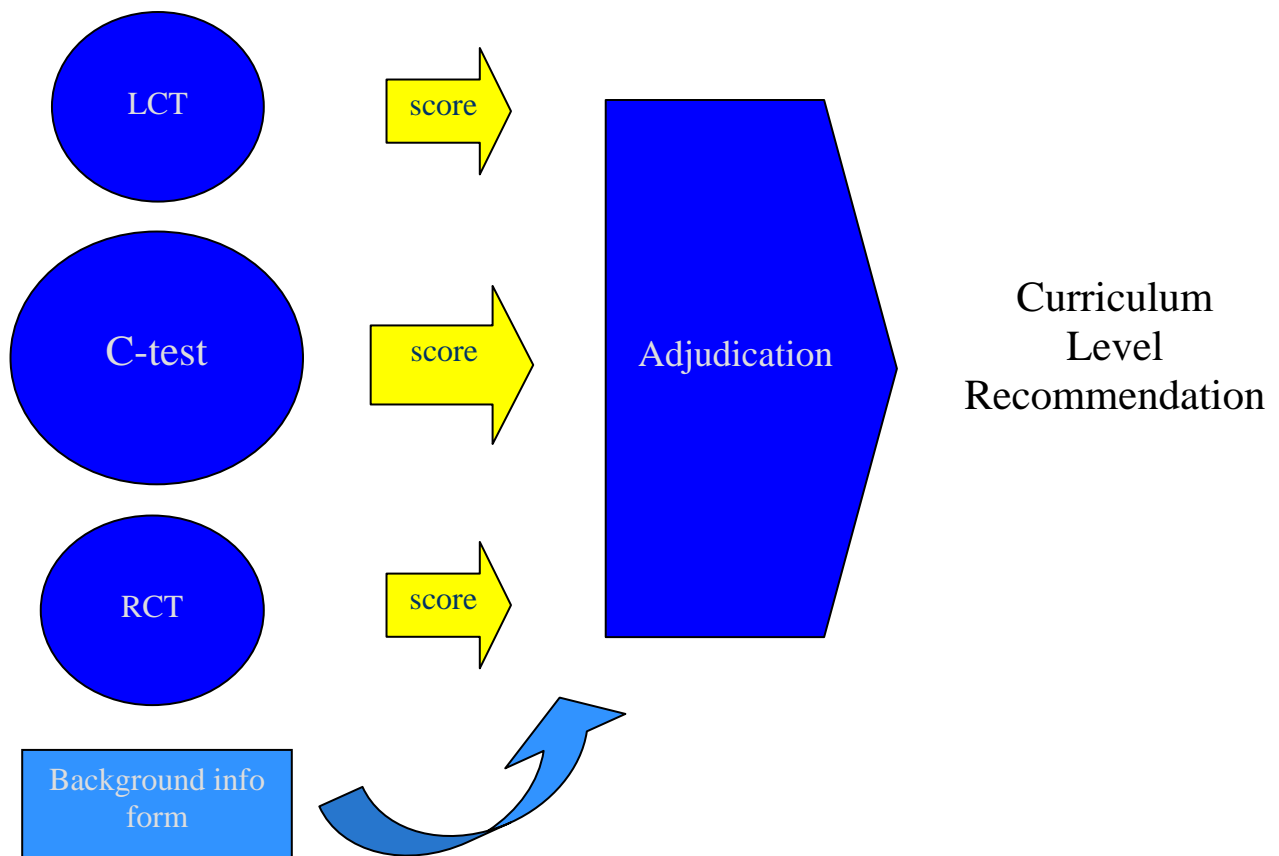
6. Implement assessment for actual intended uses

- adhere to guidelines for administration, scoring, and decision-making/use
- keep ‘field notes’ on what happens before, during, and after implementation
- seek feedback on the assessment process from those involved
- evaluate and revise assessment instruments and procedures (see section 4 below)

C. Examples: GUGD assessments

Developing and implementing the GUGD placement exam

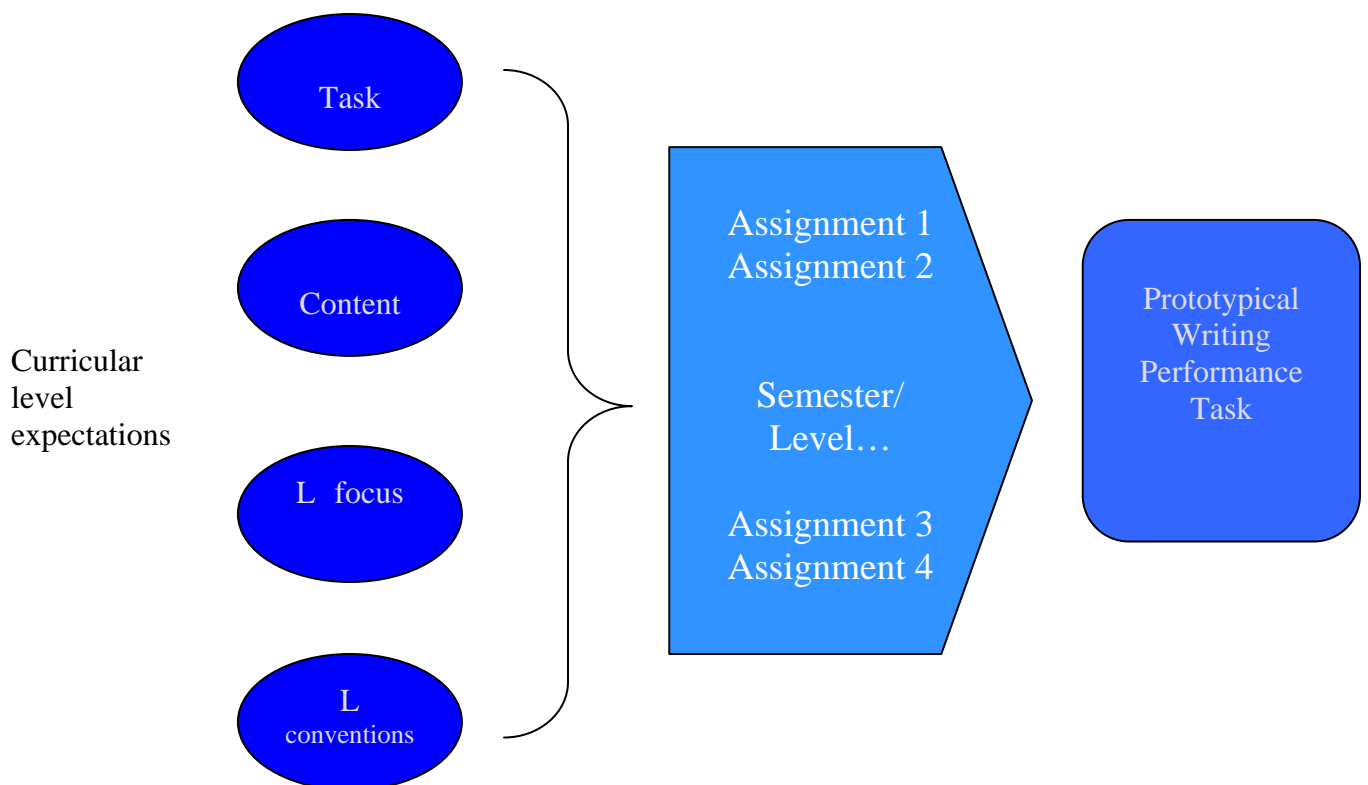
1. Prioritized qualities: curricular relevance, efficiency (immutable time and funding constraints), utility for cross-curriculum decisions, reliability
2. Review of options: precedented L2 placement procedures, GUGD curricular expectations
3. Selection of assessments: ‘best estimate’ of curricular differences within constraints



4. Operationalization: text selection by curricular-level experts, item writing, test formatting
5. Pilot-testing and implementation: considerable revisions, especially in scoring and decision-making procedures, communication with students and teachers

Developing and implementing task-based writing assessment

1. Prioritized qualities: curricular representativeness and coverage, depth of information about student writing abilities, performance elicitation consistency, utility for understanding learner development
2. Review of options: task-based assessment, GUGD curricular-level objectives and performance profiles, current writing assessment within the program
3. Selection of assessments: local curricular experts (teachers) deliberated, analyzed, and developed



Developing and implementing program-external accountability assessment???

➤ **Key References**

- Brown, J. D. (2004). Testing in language programs (revised, 2nd ed.). Upper Saddle River, NJ: Prentice Hall Regents.
- Brown, James D., & Hudson, Thom. (2002). Criterion-referenced language testing. New York: Cambridge University Press.
- Davidson, F., & Lynch, B. K. (2002). Testcraft: A teacher's guide to writing and using language test specifications. New Haven: Yale University Press.
- Norris, J. M. (2004). Validity evaluation in foreign language assessment. Unpublished doctoral dissertation. Honolulu, HI: University of Hawaii at Manoa.

4. Meeting the challenge: Validity evaluation of assessments

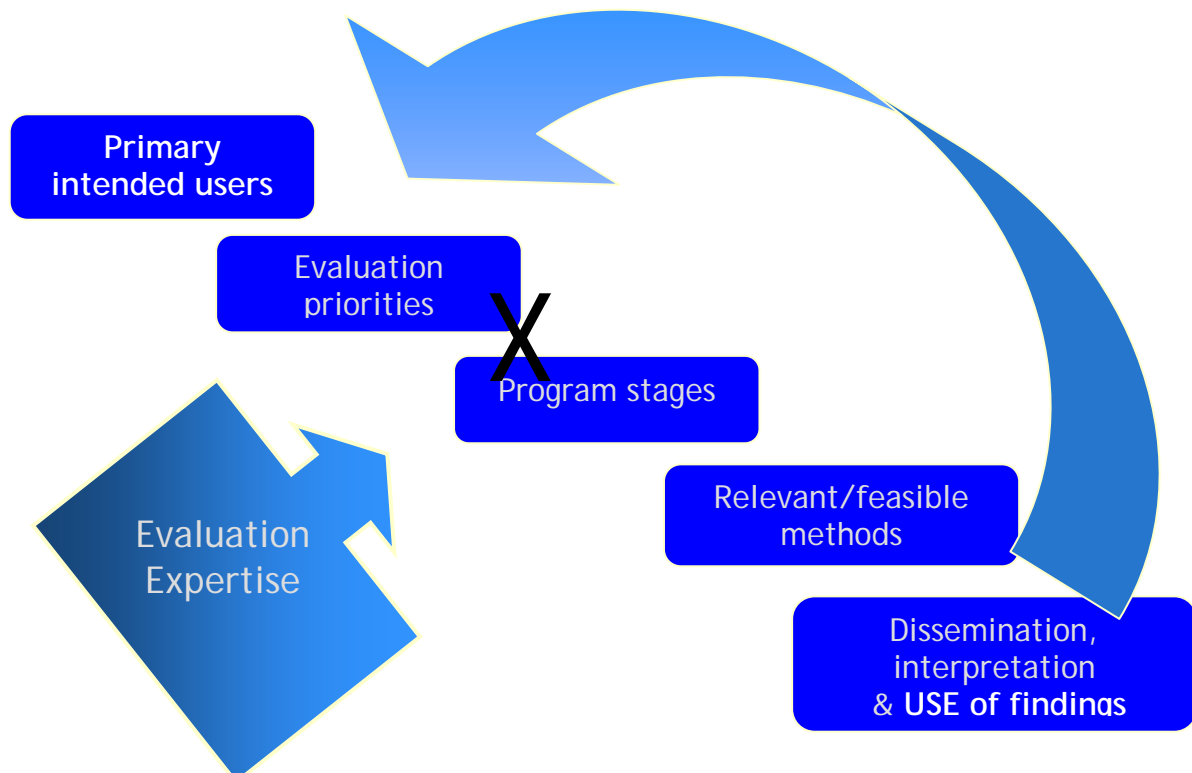
A. Rationale: Just as assessments are designed to meet specific intended uses, so too should their uses be evaluated according to these specifications. Validity does not inhere within test instruments; rather, validation is a process of evaluating particular uses for assessments with particular learners in particular contexts, in order to determine the extent to which an assessment is doing what it was intended to do.

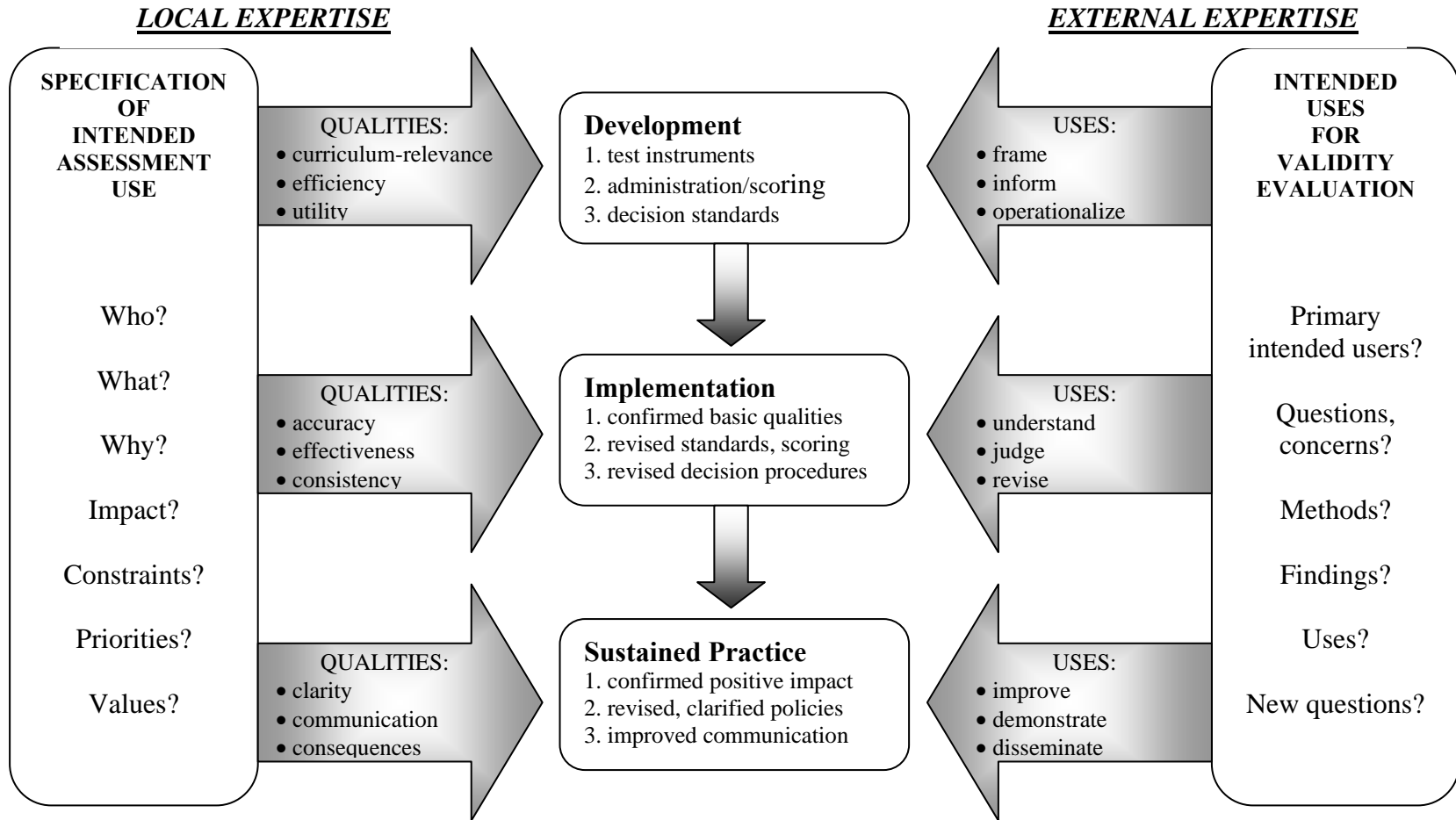
Cureton (1951): validity is an indication of “how well a test does the job it was employed to do” (p. 621).

B. Process:

- Received view: Focus on construct validity of tests as scientific measures (Messick, 1989; AERA, APA, NCME, 1999)
- Validity evaluation: Focus on provision of useful information to particular audiences for informing assessment improvement (Cronbach, 1969; Kane, 2001; Norris, 2004; Shepard, 1993)

An evaluative approach to assessment validation





C. Example: Validity evaluation schematic for the GUGD placement assessment program

A few pragmatic recommendations for validity evaluation of FL educational assessments

1. Dispense with the notion that ‘good measurement’ is all that is needed for ‘good educational assessment’; treat assessments programmatically, just like curriculum and instruction.
2. Link the ‘how’ with the ‘why’ of assessment by specifying the intended uses (who, what, why, consequences, constraints) for each assessment instance within the educational context.
3. Consider assessment development, use, and evaluation to be integrated facets of a single process which should all be informed by the clear *a priori* specification of intended use.
4. Realize that different assessment uses will require distinct assessment practices—there is no ‘one-size-fits-all’ measure of language ability (or proficiency), and assessments developed, used, and ‘validated’ for one context may not be appropriate for use in another.
5. Rationalize the purposes for evaluating assessments and prioritize the questions and concerns that make the most sense to assessment users, prior to adopting any validation methods; take seriously the expertise offered by local educators as well as the potential need for external expertise.
6. Take a cyclical approach to assessment validity evaluation within a given educational setting; prioritize which assessments are most in need of attention.
7. Enable validation to play an educationally meaningful role, rather than posing a debilitating burden, by considering the audiences for evaluation, their information demands, and the contextual constraints in articulating evaluative methods.
8. Follow through with the intended uses for evaluation, by reporting and interpreting findings, and by incorporating procedures for building consensus among assessment users on needed actions in revising, improving, or perpetuating assessments.

➤ Key References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cronbach, Lee J. (1969). Validation of educational measures, *Proceedings of the 1969 Invitational Conference on Testing Problems: Toward a theory of achievement measurement* (pp. 35-52). Princeton, NJ: Educational Testing Service.
- Kane, Michael T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Messick, Samuel J. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.
- Norris, J. M. (2004). *Validity evaluation in foreign language assessment*. Unpublished doctoral dissertation. Honolulu, HI: University of Hawaii at Manoa.
- Shepard, Lorrie A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.

5. Improving assessment capacity within college FL programs

A few ideas...

- **Take advantage of—and contribute to—language and higher education assessment conferences, organizations, journals, and other forums and resources**

Language Testing

- Main professional organization: *International Language Testing Association*
<http://www.dundee.ac.uk/languagestudies/ltest/ilta/ilta.html>
- Main professional conference: *Language Testing Research Colloquium*
- Primary journals:
 - *Language Testing* (Arnold Press) http://www.arnoldpublishers.com/journals/pages/lan_tes/02655322.htm
 - *Language Assessment Quarterly* (Lawrence Erlbaum) <http://www.erlbaum.com/journaltitles.htm>
- Overview informational web sites (see these for links to many other sites):
 - *Language Testing on the WWW* <http://www.ling.lancs.ac.uk/staff/visitors/kenji/test.htm>
 - *Resources in Language Testing* <http://www.dundee.ac.uk/languagestudies/ltest/ltr.html>
- E-mail discussion listserv: LTEST-L list; (click to join)
<http://f05n16.cac.psu.edu/cgi-bin/wa?SUBED1=ltest-l&A=1>

Higher Education Assessment

- Main professional organization: *American Association of Higher Education*
<http://www.aahe.org/initiatives/assessment.htm>
- Main professional conference: *AAHE Assessment Conference*
- Primary journals:
 - *Assessment and Evaluation in Higher Education* (Taylor & Francis)
<http://www.tandf.co.uk/journals/carfax/02602938.html>
 - *AAHE Bulletin* (American Association of Higher Education)
<http://aahebulletin.com>
 - *Peer Review* (Association of American Colleges & Universities)
<http://www.aacu.org/peerreview/index.cfm>
- Overview informational web sites (see these for links to many other sites):
 - *Internet Resources for Higher Education Outcomes Assessment*
<http://www2.acs.ncsu.edu/UPA/assmt/resource.htm>
- E-mail discussion listserv: *ASSESS* list; (to join, e-mail) listserv@lsv.uky.edu
- Example of applied assessment work in higher education:
Northern Arizona University *Office of Academic Assessment* <http://www4.nau.edu/assessment>

➤ **Pursue practical self-study in language and educational assessment**

A few excellent self-study books on *language* assessment:

Bachman, L. F. (2004). Statistics for language testing. New York: Cambridge University Press.
 Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice. Oxford: Oxford University Press.
 Brown, J. D. (2004). Testing in language programs (revised, 2nd ed.). Upper Saddle River, NJ: Prentice Hall Regents.
 Brown, James D., & Hudson, Thom. (2002). Criterion-referenced language testing. New York: Cambridge University Press.
 Davidson, F., & Lynch, B. K. (2002). Testcraft: A teacher's guide to writing and using language test specifications. New Haven: Yale University Press.

A few excellent self-study books on *educational* assessment:

Angelo, T. A., & Cross, P. K. (1993). Classroom assessment techniques: A handbook for college teachers. San Francisco: Jossey-Bass.
 Popham, W. J. (2000). Modern educational measurement: Practical guidelines for educational leaders (3rd ed.). Boston: Allyn & Bacon.
 Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. San Francisco: Jossey-Bass.

➤ **Seek formal training in language assessment (e.g., graduate degrees, professional development programs)**

A few of the primary graduate programs worldwide with an emphasis on language assessment:

University of Hawai'i at Mānoa, Department of Second Language Studies
<http://www.hawaii.edu/sls>

(Note: UH sponsors one- and two-semester graduate certificate programs for individualized professional development, including assessment)

University of California at Los Angeles, Department of Applied Linguistics
<http://www.humnet.ucla.edu/humnet/al/frames/alhome.htm>

University of Melbourne, Department of Linguistics and Applied Linguistics
<http://www.ltrc.unimelb.edu.au/>

➤ **Collaborate with assessment specialists who are working specifically on the challenges faced by college FL educators:**

Center for Advanced Study of Languages
<http://www.casl.umd.edu>

Center for Applied Linguistics

<http://www.cal.org/topics/tests.html>

National Foreign Language Center

<http://www.nflc.org>

National Foreign Language Resource Centers (central web site)

<http://nflrc.msu.edu/>

(Note: Please participate in the current national survey on placement practices in FL programs, conducted by UH NFLRC: <http://nflrc.hawaii.edu/project/placement/>)

➤ **Encourage and enable professional development of future FL educators in assessment knowledge and abilities:**

- include assessment as a core component of graduate course work
- enable participation of graduate students in all levels of assessment activities within college FL programs (e.g., development, administration, scoring, revision)
- encourage student attendance at professional assessment conferences, workshops, and other opportunities for development
- contribute to the improvement of standards for professional development (e.g., ACTFL, NCATE) by expressing your perspectives on the assessment needs of college FL programs and practitioners
- develop assessment-related initiatives within the ADFL (e.g., journal special issues, multi-site research projects)

➤ **Contact me with questions, interests, ideas, etc.:**

jnorris@hawaii.edu