

AAUSC 2014 Volume – Issues in Language Program Direction

Innovation and Accountability in Language Program Evaluation

John Norris, Georgetown University

Nicole Mills, Harvard University

Editors

Stacey Katz Bourns, Harvard University

Series Editor





**AAUSC 2014 Volume - Issues
in Language Program
Direction: Innovation and
Accountability in Language
Program Evaluation
John Norris, Nicole Mills, and
Stacey Katz Bourns**

Product Director: Beth Kramer
Product Assistant: Jacob Schott
Marketing Brand Manager:
Christine Sosa

IP Analyst: Jessica Elias
IP Project Manager: Farah Fard
Manufacturing Planner: Betsy
Donaghey

Art and Design Direction,
Production Management,
and Composition: Lumina
Datamatics, Inc.

© 2016, Cengage Learning

WCN: 01-100-101

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and
technology assistance, contact us at **Cengage Learning
Customer & Sales Support, 1-800-354-9706**

For permission to use material from this text or product,
submit all requests online at **www.cengage.com/permissions**
Further permissions questions can be e-mailed to
permissionrequest@cengage.com

Library of Congress Control Number: 2014948811

ISBN-13: 978-1-305-27509-6

Cengage Learning

20 Channel Center Street
Boston, MA
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at **international.cengage.com/region**.

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your course and learning solutions, visit **www.cengage.com**.

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**.

Instructors: Please visit **login.cengage.com** and log in to access instructor-specific resources.

Chapter 9

Do we speak the same language?: The Iterative Development of an Institutionally Mandated Foreign Language Assessment Program

Lindsay Myers and Nathan Lindsay, University of Missouri, Kansas City

Introduction: A Culture of Assessment

Developing a “culture of assessment” has become a mandate for institutions across the country (see Dunn, McCarthy, Baker, & Halonen, 2011; Suskie, 2009; Walvoord, 2010). The pressures for accountability from multiple stakeholders, and especially from accrediting bodies, are requiring colleges and universities to enhance their assessment initiatives. However, assessment initiatives at colleges and universities progress with varying levels of speed and effectiveness, depending often on the expertise, available time, and attitudes of the faculty in each department. A common challenge is to help faculty understand the benefits of assessment in the improvement of learning and teaching (Walvoord, 2010). The need for consistency and standardization across an entire institution can sometimes conflict with specific departmental goals and circumstances.

The culture of assessment nationwide and at the University of Missouri-Kansas City (UMKC) has served as an impetus for assessing student learning at the end of the foreign language (FL) major. This chapter outlines the assessment struggles and successes of a program in its nascence, and explores the relationship between university-wide requirements and the Department of Foreign Languages and Literatures’ (FLL) needs, skills, and goals. In particular, we present and reflect on the policies, practices, and structures involved in an iterative approach to assessing FL outcomes that has allowed the faculty to *close the loop*. The perspective presented here benefits from the unique collaboration of authors, one of whom serves as an administrator in the Provost’s Office and the other of whom is a faculty member in FLL. Several of the effects (both positive and negative) of a university template are outlined, including how the department is attempting to find its own voice, its own benefits, and its own meaning in this iterative process. One notable outcome of the assessment efforts focusing on outgoing majors has been the opportunity for reflection and the reinvigoration of the introductory language program. Finally, this chapter explores how FLs fit into the bigger picture of both university assessment and national trends of FL assessment. We hope that this case study can serve as a road map for other FL faculty who are facing similar temporal, financial, organizational,

or mandate-related challenges, as well as opportunities for reflection and program improvement, helping them to *speak the same language* of assessment with their colleagues and students.

The Environment of Assessment

The commitment and momentum supporting a culture of learning at UMKC are demonstrated by the number of faculty engaged in assessment and the quality of assessment plans. However, this culture has taken years to develop and has required academic departments (such as FLL) to devote significant time and energy to the assessment process. The administration's goal has been that all programs have detailed assessment plans utilizing demonstrated best practices in the student learning assessment. In addition, the administration wanted academic programs to demonstrate that they were using the student learning assessment results to modify curricula, courses, and/or policies and procedures to improve student learning. These goals and policies are outlined in the UMKC resources described in the following paragraphs.

The language of assessment is no longer foreign to most faculty, and learning outcomes are being discussed with students with greater frequency. Since the 2009 Higher Learning Commission accreditation review, UMKC has devoted significant attention to student success and retention, an effort involving extensive collaboration across campus. Such initiatives include the creation of a university college, course redesign efforts, and more intrusive advising. Assessment activities are coordinated by the Provost's Office, and the assistant vice provost for assessment has worked with faculty and staff to lead assessment plan reviews, mentoring, and communications. The assessment process, definitions, deadlines, and reports have all been standardized using a common language for goals, learning outcomes, achievement targets, and so on. An assessment template served as the initial guiding document for the creation of assessment plans for all degrees on campus (see Appendix A). The assessment plans are catalogued in a WEAVE online (WEAVE, n.d.) account, an externally hosted software application designed for assessment management.

Several assessment resources have been developed for faculty and staff during our participation in the Higher Learning Commission's Academy for Assessment of Student Learning. The assistant vice provost for assessment created a new assessment website featuring information on assessment in academic degrees, assessment in general education, assessment for the program evaluation committee, WEAVE online, and many other resources (see <http://www.umkc.edu/provost/academic-assessment/default.asp>). These resources provide a one-stop clearinghouse for faculty to learn about assessment processes. UMKC's broad commitment to assessment is conveyed in a quarterly assessment newsletter, named the "Assessable Roo," which is disseminated to faculty and staff across campus (see <http://info.umkc.edu/provost/2012/04/20/the-assessable-roo-UMKCs-outcomes-assessment-newsletter/>). Articles have featured work from a wide array of academic units on topics such as formative and summative assessment, focus groups,

survey design, and many other areas. As an additional resource, the University Assessment Committee updated the *Handbook for Learning Outcomes Assessment*, which was first developed in 2006. The handbook includes information on the core principles and processes of assessment at UMKC, the process for submitting assessment plans, an assessment glossary, a list of frequently asked questions, and several appendices (see <http://www.umkc.edu/assessment/downloads/handbook-2011.pdf>).

Addressing the Challenges Associated with Learning Outcomes Assessment

To understand the assessment obstacles faced by the FLL at UMKC, it is important to understand some of the ongoing challenges that the institution was battling at the time. Despite the various initiatives and resources established at UMKC, there were weaknesses in assessment procedures that had needed attention since 2011. Before this time, the environment at UMKC could have been appropriately described as the “Wild, Wild West of Assessment.” Many academic degrees were without assessment plans, and most faculty members were unaware of the benefits and best practices of assessment. Without a common template, there was a lack of standardized assessment language and no common reporting format. In general, as noted by our accreditation reviewers, widespread apathy and noncompliance plagued our assessment efforts. As evidenced by the lack of academic assessment plans in many departments and colleges, the institutional commitment to assessment was lacking at many different levels.

Despite these challenges, there were some positives to build upon, including faculty who were invested in teaching and intuitively engaging in assessment-like activities. Strong assessment processes existed in the health-related professional schools, and there was a culture of interdisciplinary collaboration in arts and sciences that supported our interdisciplinary iPhD program. The 2009 accreditation report indicated that the University Assessment Committee, the director of assessment, and the Institutional Research Office had all been working to improve assessment efforts across campus.

The administration dealt with the challenges mentioned earlier by creating a more faculty-centered and faculty-led initiative. For example, the administration has developed training for faculty and staff throughout the last few years. Assessment strategies and initiatives, such as the inclusion of learning outcomes in all syllabi, were addressed in New Faculty Orientation and the *Preparing Future Faculty* program, as well as in unit faculty meetings, department faculty meetings, and departmental committee structures. Assessment coordinators were identified for all programs, and assessment is the most regularly scheduled topic for workshops in the Faculty Center for Excellence in Teaching (FaCET). This organization provides ongoing support, sponsoring symposia, workshops, and webinars focused on assessment, addressing topics such as general education assessment, standardized assessments, and classroom-based assessments. Mentoring sessions are also provided by FaCET faculty fellows, as well as other faculty across campus. These mentoring sessions typically focus on the improvement of

a specific degree's assessment plan. As momentum started to grow, assessment coordinators and department chairs responded more readily to outreach and many more sessions were scheduled than in prior years. Perhaps most importantly, faculty and administrative assistants noted in their assessment reports that they put in hours of work to ensure that assessment plans were in place for each degree.

The support and promotion of assessment from top university leaders was helpful in moving our assessment initiatives forward. The provost, deans, associate deans, and department chairs were instrumental in ensuring the quality of assessment plans through their encouragement and follow-up efforts. Given the ongoing verbal support from university leaders and the assessment positions and committees in place, we are developing a sustainable infrastructure for effective learning outcomes assessment for the months and years ahead.

As a result of our efforts, assessment plans have been developed for all majors and most of the minors and certificates (which total 160 academic degrees). They have also been developed for all areas in Student Affairs and Enrollment Management, as well as for the library. By incorporating assessment plans and data in their reports, faculty have made assessment a stronger component of the Program Evaluation Committee (PEC), which reviews all academic programs on a rotating schedule every five years.

The continual outreach from the Provost's Office of Assessment to assessment coordinators, department chairs, and deans has played a particularly important role in our assessment efforts. This outreach includes offering support, answering questions, and providing training. Emphasis has been placed on developing quality rubrics, collecting data, and creating concrete action plans. The faculty mentoring model described earlier has been popular and successful among faculty members involved in the assessment process.

Despite the infusion of resources and mentoring provided, the speed and effectiveness with which various departments at UMKC have implemented these recent assessment practices varies widely. As documented in the diverse spectrum of assessment plans across campus, a department's ability to adopt best practices in assessment depends on the background, skills, and inclination of the faculty, as well as the accountability requirements that are set by other external bodies for some of the disciplines. The following sections describe the experiences of the FL faculty in balancing university assessment mandates and their own needs and interests.

Building an Assessment Program in Foreign Languages

When the university announced that all degree programs were required to create and implement an assessment program for the 2010–2011 academic year, FLL had little in place in terms of a formal assessment plan. An exception to this generalization was the existence of several documents created for a previous program evaluation review, including a mapping of the American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines onto all of our courses. This section outlines the path taken by the department to build a viable

yet ever-evolving assessment program over a three-year period, as well as the concrete steps taken to make improvements, find connections, implement the assessment plan, and analyze data from the outset.¹

One of the biggest challenges we faced during the early *Wild, Wild West* stages of the university mandate was trying to understand and fulfill the expectations of the university while balancing our department's needs, skills, and willingness to comply. This stemmed from not knowing what the expectations were, when components were due, and where to go for help. For that reason, we struggled during our first semester, reworking our plan numerous times. These early difficulties created some resistance toward the assessment process that has waxed and waned as the time commitments have grown while the benefits become clearer.

The early discussions considered the point in the program at which we should concentrate our initial assessment efforts. In the end, we selected as our theoretical foundation what Lynch (1996) refers to as the more traditional "positivistic" paradigm that focuses on quantitative, end-of-program evaluations rather than a "naturalistic," process-oriented, qualitative analysis. We felt that this approach gave us a better sense of the learning outcomes our students had achieved and was a reasonable and justifiable means to comply with the university assessment requirements.

In this climate, our early difficulties did not stem from the quality or availability of resources specific to FL assessment. In contrast, our challenge lay in extracting what would be most doable, usable, and low cost, while balancing the expectations of the university and the various language programs. As another support for our framework, Alderson (1992) gives guidelines for what he terms the "usual and logical stages in the conduct of an evaluation: planning, implementing, interpreting, reporting, using, evaluating" (p. 274). He labels planning as the "most important stage: it is difficult to adjust later for mistakes or omissions," but concedes that "perfect planning and design are both rare and unlikely" in the "real, imperfect and under-resourced world in which evaluations are usually conducted" (p. 274). It is in the context of this real world, where the stages of an assessment plan were accelerated and consolidated, that our unfunded, and so necessarily resourceful, program began.

Stakeholder Involvement

The assessment process for the three FLL bachelor degrees, French, German, and Spanish, and the common Romance languages master's degree has been dynamic and integrative, involving every faculty member in the department to some degree. The process was initially driven by external motivation and it was unclear how FLs fit the university's assessment template. To address the requirement,

¹ Reconstructing the process for the purposes of this case study was facilitated by consultation of the university-required yearly assessment narratives, which included a detailed description of the process, including an analysis of strengths and challenges that are reviewed annually by the University Assessment Committee.

FLL initially identified a single assessment coordinator for the department, and then assembled an assessment committee consisting of the chair and representatives from each language. The process also involved consultation with the faculty assessment advisor and the assistant vice provost for assessment, feedback from the University Assessment Committee, and interaction with assessment coordinators from various departments across campus. Although the administrative and organizational tasks associated with assessment lay in the hands of a small departmental committee in a small department, all full-time faculty, including the language program directors, participated in the assessment development process. For the past few years, every department meeting has dedicated part of the agenda to assessment. In short, the new commitment to assessment has necessitated the dedication of significant time and resources.

Student Learning Outcomes

The very first steps taken in developing the program involved wheel spinning and (re)-discovering what was already in place among the various degree programs. Early struggles involved trying to make the pre-existing general FLL program goals from the university catalog fit into the university-wide template (Appendix A) and its jargon. Similar to the approach taken at the University of Evansville, we sought “simplicity, manageability, and clarity in the outcomes assessment process” that would be “common to all individuals completing a major, regardless of language, area of specialization, or courses taken” (Grau Sempere, Mohn, & Pieroni, 2009, p. 142). Our process of reformulating our goals and outcomes was aided by consulting with the aforementioned university assessment faculty mentor.

The established tripartite goals fall under the major headings of linguistic proficiency, cultural knowledge, and critical thinking and analysis. These reformulated goals were derived from the departmental-wide learning outcomes that had been articulated by the chair and FLL faculty approximately five years prior and cross-checked with *ACTFL Proficiency Guidelines* (ACTFL, 2001) and *Standards for Foreign Language Learning* (ACTFL, 1996). Re-embracing these disparate yet interlinked goals has been beneficial in suggesting that we do more than cultivate linguistic proficiency. Norris (2006) highlights the unfortunately well-kept secret that “in our programs, we offer learning that is more, and much more diverse, than mere language proficiency” (p. 577). The current student learning outcomes (SLOs) are at this point in time purposefully constrained to the following: oral proficiency (SLO1), written proficiency (SLO2), cultural awareness (SLO3), and critical analysis (SLO4). For speaking and writing, students are expected to attain at least a Pre-Advanced level of proficiency as described by ACTFL in relation to their *Performance Guidelines* (1998). Although this particular proficiency designation was not selected without debate, especially concerning its inherent vagueness, the concrete descriptions of proficiency levels for language production created by ACTFL (2012) offered a common ground for discussion. In contrast, SLO3 and SLO4 were particularly challenging to define, which

we discovered during the construction and first implementations of the associated portions of an assessment rubric. We benefited from seeing our connections with other degree programs in humanities, especially History and English, when defining these SLOs and the component parts. In fact, defining the SLOs was easier when we worked backward from concrete lessons discovered in applying our tools to authentic student work. This type of simultaneous top-down (abstract to concrete) and bottom-up (concrete to abstract) approach to clarifying our definitions has been a keystone of our process, which is articulated in the subsections that follow.

Measures and Procedures

Defining our measures first involved a comparison of an outside testing program versus an in-house evaluation (see the *Foreign Language Assessment Directory* of the Center for Applied Linguistics, 2007). Some faculty members were reluctant to rely on computer- or telephone-mediated assessment, which is increasingly available even via traditional publishers, especially when not all options are available to the same degree in all three languages. The presumed built-in validity and reliability of external test appeals to the faculty. However, without available funds to pilot the tests, passing along the cost of assessment to students was not initially feasible.

In the meantime, we have chosen to use two in-house measures that were already partially in place within the degree programs. An exit interview was already being used for French, and so this was adopted as a common practice to measure the oral proficiency for German and Spanish as well. Two full-time faculty members from each language participate in each interview. This decision was fairly unproblematic for German, but created a rather significant burden for Spanish due to the number of graduates in any particular year. The Oral Proficiency Interview is currently a department-designed interview relying on the faculty's pedagogical training without having to rely on a particular standardized test. We have constructed oral interview guidelines with suggested topics related to the individual student's experience at UMKC (see Appendix B). For this purpose, although imperfect, the interview becomes a means to gather qualitative data about the student experience while assessing spoken language skills.

Written proficiency, cultural awareness, and critical analysis are measured by evaluation of a graduation requirement for all three languages: the capstone paper. Having all three of these SLOs evaluated simultaneously in a single rubric offered us, as Byrnes, Maxim, and Norris (2010) describe, a "means for reminding both teachers and learners that writing performance is not merely about creating good sentences and paragraphs or using accurate grammar" (p. 154). Using it as a framework for SLO evaluation offered us the opportunity to see how we were implementing this capstone requirement in different ways. Creating a single rubric to reflect the capstone expectations and experience for each of the languages and the instructors required extensive discussion and reflection. An overview of the selection, modification, and implementation of rubrics follows.

Rubrics

ACTFL's (2003) *Integrated Performance Assessment* served as a starting point for our purposes. We adopted the three-tier evaluation system with the performance levels of "Exceeds," "Meets," or "Does Not Meet" for its simplicity. For the Bachelor of Arts (BA), we use two rubrics: one for oral proficiency adapted from the original Interpersonal Rubric for the Pre-Advanced Learner and the other adapted from the original Presentational Rubric for the Pre-Advanced Learner for the three SLOs measured in the capstone Paper.

For the most part, the oral proficiency rubric remains rather consistent with the version created by ACTFL intended for assessment of the interpersonal mode of communication. We made some modifications to the ACTFL oral performance rubric, such as word choice changes and category additions. For example, significant modifications resulted from post-pilot year FLL faculty feedback, suggesting that the category of "Language Control" was too broad. Therefore, "Fluency" was given its own category, as was "Complexity" after several iterations (see Appendix C). The need for this type of modification came from the observation that, in rating student performances, faculty members were identifying two levels simultaneously (e.g., "Meets" and "Does Not Meet" for a single criterion). This indicated that our distinctions between two evaluation levels were unclear or that the category needed to be broken down into separate categories.

By contrast, the modifications necessary to create a rubric encompassing three SLOs were more significant. Almost all of the categories evaluating SLO4 (critical analysis) were additions to the ACTFL rubric (see Appendix D). Such additions (with thanks to our colleagues in History and English for sharing their relevant rubrics) included "Research," "Overall Structure and Organization," and "Citation Format," except for parts of "Impact," whose wording we found to be consistent with our assessment needs in critical analysis. The two categories chosen to evaluate SLO3 (culture) were additions as well. A remarkably tricky area for assessment has proven to be defining the essential components of "Cultural Awareness" evaluated in the capstone Paper. After the first evaluation cycle, we noticed a few significant score discrepancies between different evaluators for the same criterion. Much attention was given to normalizing our instrument throughout the third academic year in faculty meetings. This focus aligns with Webster Goodwin's (2011) view that "[r]ubrics, like our goals for students' learning and our assessments, must be conceived as an ongoing dialogue (explicit and implicit), and are themselves subject to evaluation" (p. 135). The procedure involved dedicating a portion of each faculty meeting to a particular rubric or subset of the rubric. Led by the FLL assessment coordinator, all full-time FLL faculty talked through each criterion and its label, and pointed out obvious inconsistencies from the data collection. For example, they discussed why the results of one language were lower than the results of another language. Was this due to faculty grading habits and expectations? Or do they reflect the reality of the student population? Although the process itself seemed tedious at times, it has resulted in improved buy-in as well as greater confidence in both validity and reliability of the instruments (see the section Improving the Process).

Data Collection

The process of collecting and analyzing data has slowly evolved, although it remains one of the more challenging aspects of our plan. The first cycle involved significant backward scrambling since our measures were defined and rubrics honed around the same time the students were graduating. The small numbers of that first cycle reflect our difficulty in tracking down papers that may have been, in some cases, submitted a semester earlier. Each language division was charged with organizing the evaluations, and then they submitted the completed rubrics over the summer to the FLL assessment coordinator, who hurriedly attempted to meet the university WEAVE reporting deadline. In the second cycle, the language divisions were aware of the assessment expectations, although a lack of program-internal deadlines still culminated in summer-end scrambling within all three divisions and on the part of the FLL assessment coordinator, who consolidated, entered, and analyzed all the data. The third cycle marked a significant turnaround for data collection as each language section created a flowchart and a data repository. Each FLL Assessment Committee member served as the organizer for (increased) faculty involvement and the repository for completed rubrics. The data analysis was initiated and summarized by the FLL assessment coordinator but involved the entire committee.

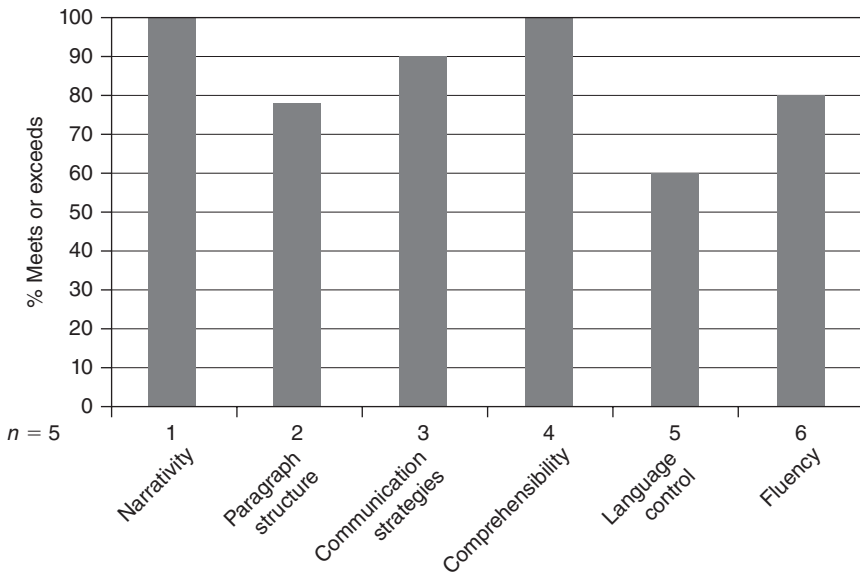
Initial Outcomes

Our first set of results told us much more about the assessment process and our rubrics than it told us about our students. With only 11 students being assessed in the first cycle, the results were not substantial or representative. Although theoretically sound, the university requirement to define achievement targets early in the development of the assessment plan led to some rather time-intensive and uninformative number crunching. Focusing on the required definitions of success, we concluded our first assessment cycle by reporting on the percentage of students that had attained “Meets Expectations” or “Exceeds Expectations” in a majority of categories on the rubric (with the benchmark at 80%), which was not useful for a conversation about student language learning. Notably, in the first cycle of reporting, we recorded that the French students had not met the prescribed definition of success. Only 75% of their students were attaining “Meets Expectations” or “Exceeds Expectations” in a majority of categories on the speaking rubric. However, in reality, this meant that 3 out of 4 assessed students had met or exceeded our expectations. Furthermore, it was not readily apparent where, within the subcategories of the speaking rubric, the students had received designations of “Does Not Meet,” “Meets,” and “Exceeds” expectations. The limitations of the initial results and our analytical approach effectively discouraged meaningful discussion, especially since the particular strengths and weaknesses were unrecoverable in this form. In essence, the departmental assessment committee learned that the university WEAVE template requirements were not necessarily commensurate with FLL faculty expectations for meaningful and interpretable results of the outcomes in linguistic proficiency, cultural knowledge, and critical thinking and analysis.

In contrast, at the end of the second assessment cycle, the FLL assessment committee focused on making the most of the results so that FLL faculty could appreciate the outcomes of assessment efforts by presenting them in a transparent fashion that would facilitate a worthwhile discussion, so that FLL faculty could appreciate the outcomes of assessment efforts. Instead of considering holistic targets and success rates, as we had done after the first cycle to fulfill our WEAVE obligations, the results were broken down according to the rubric subcategories. For each language and SLO, a bar chart, like the one displayed in Figure 9.1, was created, which allowed the department to observe, respond to, and potentially refute trends in the data. As an illustration, with the understanding that the insufficient numbers prevented steadfast conclusions, the faculty as a whole could easily interpret that the strengths in speaking proficiency were “Narrativity” and “Comprehensibility” and the clear relative weakness was in “Language Control.”

Indeed, one of the primary observations made from discussing and analyzing the resultant tables was that across languages and across modes of expression, the category of “Language Control” consistently lagged behind the other subcategories. For instance, in speaking proficiency, 70% of Spanish students received “Meets or Exceeds” evaluations in “Language Control,” whereas 95% of Spanish students received “Meets or Exceeds” evaluations in “Comprehensibility.” Similarly, “Language Control,” at 87%, was the only subcategory of writing proficiency that did not attain 100% combined “Meets or Exceeds” expectations.

Figure 9.1 Sample Second-Cycle Speaking Proficiency Outcome.



(The reader may note the discrepancy between the categories presented here and those described earlier in text. The “Complexity” category was not added until the third cycle.)

The “Language Control” results are consistent with the profile of a Pre-Advanced language learner who, according to the ACTFL Performance Guidelines (Swender & Duncan, 1998), is “most accurate when narrating and describing in connected sentences and paragraphs in present time with decreasing accuracy in past and future times” and “may continue to exhibit inaccuracies as the amount and complexity of language increases” (p. 490).

The third cycle marked a significant breakthrough for our data collection since for the first time, the number of students assessed (39 students, or approximately 90% of the graduates in our department) allowed us to make more assertive claims and conclusions and articulate a more representative success rate. Increasing numbers also offered us the opportunity to reflect on previous trends and confirm or deny preliminary reports. For example, we found that the speaking proficiency category with the fewest “Does Not Meet” evaluations was “Comprehensibility” (6%) and “Language Control” (17.5%), and “Fluency” (17.5%) received the most “Does Not Meet” evaluations.

This pattern was echoed in the writing proficiency results, for which the category with the most “Does Not Meet” evaluations was again “Language Control” (14%). Overall, more students met our definitions of success in writing than in speaking, potentially due to the respective planned versus spontaneous nature of the measures. The third cycle results confirmed the overall strength in “Cultural Awareness” that had become apparent in the second assessment cycle, whereas it had been masked by poor category articulation in the first cycle. Its subcategory “Cultural Content” received “Exceeds” designations in 50% of the evaluations. In contrast, some of the critical analysis results pointed to continuing problems in rubric interpretation. For example, nearly 13% of the “Citation” evaluations involved cases in which one instructor selected “Exceeds Expectations” and the other selected “Does Not Meet Expectations.” Further, it was noteworthy that despite an overall pass rate, the subcategory “Depth” received fewer “Exceeds” designations (18%) than any of its counterparts in critical analysis.

Looking back, the FLL assessment team recognizes that nearly all our preliminary results have become useful indicators for either assessment improvement or the capability to discuss and articulate the emerging strengths and weaknesses of our program and of our students. By capitalizing on assessment processes already in place, and documenting results so early in the development of the plan itself, we accelerated the dynamic loop of implementation, reflection, and programmatic change. These insights would not have been available after just one or two assessment cycles, and we find that this longitudinal approach is a key component to keep in mind during early-stage assessment efforts.

Improving the Process

It was clear that the shift to a focus on results that were more relevant to our department’s interests rather than the requirements of the online assessment template in WEAVE led to increased interest in the process itself. The second cycle concluded with clear action plans for the third cycle that allowed us to better

interpret findings and spend more time on student learning than on the act of assessment. For example, as mentioned earlier in text, our first set of results indicated a lack of inter-rater reliability. Notably, in French, for the first evaluation of SLO3 “Culture,” not one student received the same designation from the two evaluators. In contrast, in the second cycle, French reviewers agreed on cultural evaluations 60% of the time. Moreover, as outlined earlier, our biggest challenges in the second and third years were related to our internal processes of organizing data collection and enforcing deadlines. These included issues of document management, assessment assignments, and timing. The FLL department discovered that a more articulated process with clear, delegated tasks within every language section was needed in order to spend assessment time reflecting on student learning outcomes, rather than urging faculty to submit their evaluations before deadlines.

The timing of the assessment is also a challenge. In general, most of these assessments take place during the last week of class or during finals week. This is particularly problematic since it falls at the end of the semester, already a busy time for assessment. Additionally, although we benefit from the increased reliability of using multiple raters for any student production (whether oral or written), this component is organizationally taxing. It was also not immediately clear how to best keep track of graduating students. When is the best time to assess a student? When they fill out graduation paperwork? Or when they enroll in the capstone Course? How does the assessment team know who is graduating? The total number of students assessed in FLL was smaller than we would have liked because of a lack of unified answers to the previous questions.

Many of the internal process challenges were resolved with more efficient and organized document management. The department has taken advantage of network shared files and course management systems to create repositories for student work, assessment evaluation assignments, as well as rubrics and results. As a result of our organizational initiatives, for the first time, nearly all outgoing students in all languages were assessed in the third cycle, which created the potential for more robust data sets.

Another timing issue has resulted from the university assessment requirements. There is a WEAVE October 1 deadline every year that creates an *assessment crisis* period in early fall. Our challenge has been to limit the amount of time that is required to do the reporting in WEAVE in order to spend that energy on program improvements. Our efforts to create and interpret more meaningful data also created a statistical challenge for our department, which relies on outside advising or skills in statistical analysis and report formatting.

As we anticipate our fourth assessment cycle and beyond, we are eager to benefit from all of the time-consuming attention that we paid to the process and begin to focus on student-centered and programmatic action plans. We also now have a specific blueprint to embark on a meta-evaluation of the quality of the assessments themselves, as advocated by Norris (2008). In other words, now that FLL has the requisite and functional mechanics in place, and has fulfilled the requirements of the University Assessment Committee, the next self-imposed

chapter demands a rigorous phase of reflection on the validity of the assessments themselves.

Benefits of the Assessment Program(s)

As noted earlier in text, the university has demonstrated improved assessments and the documentation of assessment practices for accreditation purposes. Despite the challenges seen in the first three years of the FLL assessment program, which often involved balancing university requirements and deadlines with departmental priorities, there were immediate and clear benefits. Although faculty sometimes asked whether or not we learned anything that we did not already know, there were many advantages to this process. These advantages include awareness of assessment terminology and process, student learning, and various department benefits.

For example, faculty members and students in FLL have seen several advantages in becoming more aware of assessment terminology and processes. These types of changes have created a more unified and consistent approach to assessment in our program. As a result, students are exposed to program goals and the associated language from the beginning of language study, making the program-end measures more cohesive and less jarring. Such intentionality and transparency has become a model for other academic programs at UMKC.

Furthermore, based on the minimal results in our first two years of data collection, even our anecdotal and qualitative observations inspired immediate curricular changes. FLL has begun the process of normalizing the basic elements of assessment (learning outcomes, rubrics, etc.) into courses and programs. The course directors in French, German, and Spanish have reported that they use and adapt the program-final rubrics according to their course and level so that students are introduced to the jargon and expected outcomes earlier in the program. This integration is particularly visible in the Spanish syllabi that now include language seen in the program-final rubrics. The Spanish assessment representative, in conjunction with the course directors, rearticulated formal and informal oral production outcomes by customizing the department's rubric for evaluating speaking in the Novice levels. These modified rubrics are currently distributed with the syllabus in all sections of first-year Spanish. Early results from two measures, the oral interview and the capstone Paper, prompted significant reforms. First, the observed weaknesses in oral proficiency led to an increased focus on this skill throughout the major, rather than relegating it to the early language sequence. In French, for example, simulations of the oral interview experience now take place at least once a year in required major courses. A required course of the major at the 300-level was also modified to focus on speaking strategies. Second, capstone Course syllabi were restructured to create more time for attention to form following the department's hypotheses that the capstone Papers were not sufficiently reflecting student abilities. These initiatives were examples of the department's attempt to close the loop or use assessment data to make meaningful and specific changes (see Walvoord, 2010).

Significant gains were also seen in the realms of program evaluation and restructuring. After a completed year of pilot assessment, the department had its five-year external program review. During this same year, we began the process of merging our three BA degrees into a single BA degree. In developing a single, shared assessment program for one degree, it allowed for cross-language conversations on practices and conventions, and many of the requisite conversations for the conversion were expedited.

The creation of a single BA degree and accompanying student assessment plans also facilitated the articulation of an assessment plan for the shared Master of Arts (MA) program in Romance languages with an emphasis in either French or Spanish. Writing both assessment plans consecutively cemented and clarified connections between the BA programs and the MA program, and highlighted the flow of expectations from one degree to the next. For example, after the expected outcomes for the pre-Advanced proficiency level were designated as “Meeting Expectations” for the BA degree in both speaking and writing, the designation “At Least Advanced” proficiency was adopted for the MA degree. Furthermore, formulating the requirements for the MA program occurred at the same time that the FLL faculty was rewriting degree requirements, and therefore, the two processes became productively intertwined. This development was especially apparent as the faculty of the two emphases created more cross-language consistency in terms of the examination process on which the assessment plan was primarily based. Specifically, the program committee for the MA degree with emphasis in French added a written exam component, previously found only in Spanish, and the MA degree with emphasis in Spanish adopted a speaking exam component.

One of the most positive byproducts of our assessment efforts in the last two years of the assessment process was the meta-reflection and communication encouraged among faculty. Similar to Saxton and Mance’s (2011) assessment discussions in their English department, we also found that “our assessment-driven discussions about our differences of opinion and our preferences, our particular specialties and styles have created a more intellectually exciting and cohesive department at the level of both curriculum and community” (p. 273). For instance, there has been explicit effort to increase the frequency of course rotation, syllabus sharing, and scope and sequence discussions. Likewise, we engaged in constructive discussions about academic writing proficiencies, the definition of culture, and the strengths and weaknesses of our students. It also gave us the opportunity to communicate and articulate our similarities and differences across the department as a whole. For example, looking together at the assessment results of capstone Papers in the three languages has engendered conversations about how the capstone Paper assignment is presented, implemented, and revised in each division. Several notable differences, many of which result from the inherent differences in program size, include the timing and scope of the capstone Courses. Whereas the German division treats each capstone Project as an independent study that sometimes stretches over an entire academic year, the Spanish capstone Course, offered in the fall, often has 30 students and the French version typically involves 10 students in or near their last semester of study.

An additional benefit of the nascent FLL assessment program has been increased visibility on campus. Our success in developing a coherent, functioning program in a short amount of time has presented opportunities to talk about the program development to a wider university audience. As Norris (2006) highlighted, “specification of the outcomes of college FL programs provides us with a decisive opportunity to state who we are, why we exist, what our value is to learners, institutions, and society, and quite frankly, why we should not be shut down” (p. 577). To this end, we have been able to share successes and lessons learned at symposiums and in newsletters. For instance, the FLL assessment coordinator has presented at faculty workshops on assessment organized by the assistant vice provost of assessment, as well as a campus-wide faculty symposium. She was also one of the first faculty contributors to the aforementioned “Assessable Roo.” Through this exposure, we have found commonalities in assessment among disciplines (e.g., between assessing language and medicine). By sharing its best practices, the FLL has also played a role in promoting good assessment practices at UMKC.

Sustaining Momentum and Looking Forward

Much of the third cycle of assessment was dedicated to improving the process and getting better results for the amount of work required, with the goal of making assessment more of a logical offshoot of what we do. As noted earlier, this involves moving from a defensive reaction to top-down, mandated assessment processes to greater departmental ownership in the process. Integrating Walvoord’s (2011) assessment components, FLL has created an assessment system that is undoubtedly “consonant with the requirements of the accreditation bodies” and “consonant with the culture and values of the department” (p. 335). It is also helpful to students and faculty and “manageable in terms of time, resources, expertise, effort, and work load” (p. 335).

To that end, the FLL faculty is considering where our energy is best spent to help us “speak the same language” of assessment. Is it reasonable to continue to conduct all assessments ourselves? The affective and qualitative benefits of conducting our own oral assessments are ample, yet the time commitment is appreciable. Without further training, mapping to ACTFL proficiency levels is problematic with respect to validity. Outsourcing the language proficiency components of our assessment would allow us to more confidently compare results across languages as well as potentially add all four skills.

Furthermore, the department began with a summative assessment plan, but a logical extension involves reasserting and eventually assessing those critical benchmarks with more formative methods. One approach to more formative methods is the use of the biographical and demographic data collected during assessment to learn more about the sources of student success. It is clear that not all FLL graduates have had the same experience. Factors such as study abroad, heritage language exposure, and where the student entered the program (often related to years of high-school exposure) all have a significant impact on student performance.

In this way, a simultaneous top-down and bottom-up view of assessment may enhance both the outgoing results and the achievement of minors and other students who take courses in the department. The connection of these

efforts to Language Program Coordination has become clear and advantageous at our institution. We are currently extending the assessment program to the elementary and intermediate language courses as an essential component of the recently reaffirmed FL requirement in the college. The successful and enhanced cross-language communication resulting from the program-end evaluation supports our efforts to mirror the articulation of learning outcomes, measures, and assessment at the introductory and intermediate levels. Moreover, since all full-time faculty have been involved with the program-end assessment program from the beginning, the work of extending it to the language program and part-time instructors is logical and conventionalized. Since the work of creating a culture of assessment has been established in our department, the language program directors are currently benefiting from the increased time and energy devoted to reflection on skills, approaches, and best practices.

As Bernhardt (2006) points out, “focusing on student achievements has reverberative effects on professional development in its many facets as well as on the perception of the efficacy of language programs throughout an institution” (p. 588). This focus on student learning outcomes in the FL programs has certainly given us and will continue to give us the opportunity to enhance the perception of FLL at UMKC.

Conclusion

The assessment program that we developed at a rapid pace in response to the requirements of a university-wide mandate has some distinctive features. Foremost among these resourceful and dynamic aspects is the iterative, longitudinal process that has led to stronger rubrics and more specific action plans. We have tried to paint an accurate picture of the benefits and challenges of developing such a program. We recognize that each institution and department has its own peculiarities and obstacles, but we hope that many of the approaches and resources described throughout the chapter can serve as a guide to those who are initiating this process. At the very least, we have sought to demonstrate that even nascent and imperfect assessment plans can herald change at many levels.

As noted earlier, there have been many advantages of the mandated assessment plan and template for FLL. The university requirements resulted in a full-fledged, functioning assessment program in a very short amount of time. The requirements of the process encouraged discussions among faculty on topics that might not otherwise leave the classroom. Moreover, the culture of assessment at the end-of-program level has led to a culture of reflection on best practices that has accompanied renewed interest in the introductory language sequence.

The disadvantages of mandates and required templates for FLL should also be acknowledged. A hurried environment does not always lead to the best possible product. It has been a difficult challenge to transition from the mindset that we are fulfilling a requirement to a mindset of commitment and ownership.

We have also learned some important lessons in how to make the time-consuming requirements of compliancy remain meaningful to the department's needs. We now recognize how to “do less assessment” and more work based on its outcomes. By focusing on student learning, we have been able to engage

the faculty more effectively in this process. By iteratively examining what our students really know at every level of our program, we are better able to communicate with one another, and we have a better understanding of whether we *speak the same language* as we work to improve teaching and learning in foreign languages.

Appendix A

UMKC Assessment Plan Template

Assessment Cycle: 2012-2013	Date Submitted:
Academic Degree Program:	Degree Level:
Department/Division:	Submitted by:
	Contact Email:
	Contact Phone:
	College/School:

Mission (Mission of the Academic Degree Program)

The academic degree program's mission statement links the program to the department, college/school, and ultimately to the overall mission of UMKC. *(Does your academic degree program mission support the UMKC mission, your college/school mission, and your department mission?)*

Goals

A goal is a general statement about the aims or purposes of the educational experience in the academic degree program. Goals are long range outcomes that are written in broad language.

Goal 1:

Goal 2:

Goal 3:

Outcomes/Student Learning Outcomes

An outcome is a specific statement that describes a desired learning outcome for the academic degree program. *(At least 3-5 student learning outcomes should be identified for each academic degree program; Learning outcomes should be specific, measurable, ambitious, and time bound; Each student learning outcome will be associated with a goal.)*

SLO1

SLO2

SLO3

SLO4

SLO5

Measures

*Identify and describe appropriate assessment instruments or methods for each student learning outcome and establish a criteria level for learning success (achievement target/success criteria). *(Multiple measures including direct/*

indirect, formative/summative are desirable for each student learning outcome; Collect sources of evidence that will convince you and others that your students are reaching the desired learning objectives.)

SLO1

SLO2

SLO3

SLO4

SLO5

Findings

Analyze the findings from your assessments. Come to a clear understanding and agreement on areas that still present opportunities for academic degree program growth and improvement. *(What are the implications of the findings? Are students achieving expected learning outcomes? What program changes could you make to improve student knowledge and skills that did not reach criterion success levels? What can you infer from the data?)*

****Document the findings of assessment. Summarize the results for reporting purposes; be sure to retain detailed documentation on file for reference purposes if needed (accreditation, program evaluation, etc).**

SLO1

SLO2

SLO3

SLO4

SLO5

Action Plan

Strategies for using results for program improvement development, methods for reporting results, timeline and responsibility for assessment activities.

Part 1: Reviewing student learning outcome data and making adjustments to the academic program. *(What future actions should your program take? How can you assist students develop the learning outcomes you wish them to achieve?)*

Click here to enter text.

Part 2: Reviewing and making adjustments to the academic assessment plan. *(Are changes necessary in your objectives? Are your assessment methods providing you the quality and quantity of information you need?)*

Click here to enter text.

*Be sure to identify & include the State mandated exit exam your unit has declared (major field test or other identified exit competency exam).

** The quality enhancement process is continuous and includes completion of annual assessment cycles that use the results to make improvements to your academic program. Improvements might include revising organizational structure, reallocating resources, revising administrative policies/procedures, revising curriculum, individual course revision, sequencing of courses, inclusion and/or modification of educational experiences and strategies (e.g., undergraduate research, internships, practicum, study abroad, service learning).

Glossary of Terms

Achievement Target/Success Criteria: overall level for satisfactory performance on a student learning outcome

Action Plan: activity sequence designed to help accomplish intended outcomes/student learning outcomes and/or improvement of academic assessment plan

Findings: assessment results for comparison of actual vs. expected achievement level

Goal: broad statement about desired ends

Measure: method to gauge achievement of expected results

Mission: highest aims, intentions, and activities of the entity

Outcome/Student Learning Outcome: measurable statement that describes the knowledge, skill or ability students will possess upon achievement of that outcome as it relates to the mission

Developed utilizing & modifying the following documents:

Virginia Commonwealth University - <http://www.vcu.edu/quality/pdfs/WEAVEManual2002.pdf>

University of Western Washington - http://www.wwu.edu/depts/assess/prog_handbook.pdf

Western Association of Schools & Colleges - http://www.wascsenior.org/findit/files/forms/Program_Learning_Outcomes_Rubric_4_08.pdf

Appendix B

Oral Interview Guidelines

Duration: Approximately 25 minutes

Warm Up:

Simple Background Questions

Current Exam Schedule

Level-Checks & Probes:

Where do you live? Where have you lived during your time at UMKC? Describe one of those places... How do you get to campus? What route do you take?

What are you going to do next year? this summer/winter break?

Why did you choose UMKC? What did UMKC offer you? the major?

Do you remember your first X language class? Where was it? What was your reaction? And if applicable, your first X language class at UMKC?

Have you studied abroad? What did you learn? How was that experience? If not, do you want to? Where (else) would you like to go?

What has been the best part of the language major experience? the most challenging? What is the most memorable event at UMKC? Describe it...

What would you change about your experience? What recommendations would you give to other majors?

What role do you think X language will have in your life? What are your career plans?

Role Play Topics:

Pretend we are skeptical parents of prospective language majors at UMKC. You are giving them a tour of campus and telling them about the major.

Why is it important to study X? What do you say to people who ask you why you majored in X?

Pretend you are abroad and are talking to a group of people who think that all students at American universities are X (=only interested in partying, insular, loud, etc.) and ask your opinion. Is it true?

Wind Down:

(How) are you celebrating graduation? What are your plans for the rest of the day?

Appendix C

Additions to ACTFL's (2003) "Interpersonal Mode Rubric Pre-Advanced Learner" for Oral Proficiency Evaluation

Category	Exceeds Expectations	Meets Expectations	Does Not Meet Expectations
Complexity of lexical and grammatical structure	Successful use of complex grammatical structures <u>and</u> advanced vocabulary.	Some successful attempts at complexity in lexical and/or grammatical structure.	Primarily uses elementary structures and vocabulary.
Fluency	Expression is highly fluid and smooth.	Language production is mostly fluid and smooth.	Language is slow or halted.

Appendix D

Additions to ACTFL's (2003) "Presentational Mode Rubric Pre-Advanced Learner" for Capstone Paper evaluation

	Exceeds Expectations	Meets Expectations	Does Not Meet Expectations
Cultural appropriateness of the language SLO2	Clear evidence of culturally appropriate language.	Some evidence of culturally appropriate language.	Little or no evidence of culturally appropriate language.
Cultural Awareness & Analysis SLO3	An analysis and reflection on cultural products and/or practices and differences within.	A description supported by factual information of cultural products and/or practices.	A superficial description of cultural products and/or practices often limited to stereotypes.
Cultural Content SLO3	Displays deep knowledge and incorporates context underlining the complexity of the topic.	Displays accurate and relevant (historical) contextualization.	Displays lack of historical depth or erroneous, superficial or deficient information.

Continued

	Exceeds Expectations	Meets Expectations	Does Not Meet Expectations
Overall Structure and Organization. SLO4	Paper is very clearly structured, follows a logical development of ideas and is very well organized. Excellent introduction: thesis clearly stated, engages the reader. Excellent conclusion: restates and develops the thesis, draws together all strands of the argument, not redundant.	Paper is clearly structured, and mostly follows a logical development of ideas. It is reasonably well organized. Good introduction: thesis stated, if implicitly, somewhat engages the reader. Good conclusion: restates the thesis, draws together some strands of the argument.	Paper is poorly structured, and rarely if ever follows a logical development of ideas. It is not well organized. Poor or no introduction: no thesis stated or poorly stated, does not engage the reader. Poor or no conclusion: does not restate the thesis, draws together few if any strands of the argument.
Depth of writing and analysis SLO4	Describes, analyzes and critically evaluates multiple topics in support of the thesis.	Describes and analyzes multiple topics in support of the thesis.	Insufficient analysis or breadth of supporting topics.
Research SLO4	Utilizes a variety of appropriate sources and includes critical thinking about findings.	Utilizes appropriate sources and includes findings from research.	Includes little or no findings from research and/or utilizes inappropriate sources.
Citation Format SLO4	Citation format is consistent and accurate.	Citation format is generally consistent and accurate.	Citation format is inconsistent or inaccurate.

References

- Alderson, J. C. (1992). Guidelines for the evaluation of language education. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 274–304). Cambridge: Cambridge University Press.
- American council on the teaching of foreign languages. (1996). *Standards for Foreign Language Learning: Preparing for the 21st Century*. Hastings-on-Hudson, NY: Author.
- American council on the teaching of foreign languages. (2001, 2012). *ACTFL Proficiency Guidelines*. Hastings-on-Hudson, NY: Author.
- American council on the teaching of foreign languages. (2003). *ACTFL Integrated Performance Assessment*. Hastings-on-Hudson, NY: Author.
- Bernhardt, E. (2006). Student learning outcomes as professional development and public relations. *The Modern Language Journal*, 90(4), 588–590.
- Byrnes, H., Maxim, H., & Norris, J. (2010). Realizing advanced L2 writing development in a collegiate curriculum: Curricular design, pedagogy, assessment. *Modern Language Journal*, 94. Supplement to vol. 94, monograph.
- Center for Applied Linguistics. (2007). *Foreign language assessment directory*. Retrieved August 30, 2013, from <http://www.cal.org/CALWebDB/FLAD>
- Dunn, D. S., McCarthy, M. A., Baker, S., & Halonen, J. S. (2011). *Using quality benchmarks for assessing and developing undergraduate programs*. San Francisco, CA: Jossey-Bass.

- Goodwin, S. W. (2011). Fearful symmetries: Rubrics and assessments. In D. Heiland & L. J. Rosenthal (Eds.), *Literary study, measurement, and the sublime: Disciplinary assessments* (pp. 133–151). New York: The Teagle Foundation.
- Grau Sempere, A., Mohn, M. C., & Pieroni, R. (2009). Improving education effectiveness and promoting internal and external information-sharing through student learning outcomes assessment. In J. Norris, J. Davis, C. Sinicrope, & Y. Watanabe (Eds.), *Toward useful program evaluation in college foreign language education* (pp. 139–162). Honolulu: National Foreign Language Resource Center.
- Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge: Cambridge University Press.
- Norris, J. (2006). The why (and how) of assessing student learning outcomes in college language programs. *The Modern Language Journal*, 90(4), 590–597.
- Norris, J. (2008). *Validity evaluation in foreign language assessment*. New York: Peter Lang.
- Saxton, K. T., & Mance, A. J. (2011). From skepticism to measured enthusiasm: The story of two literary scholars' introduction to assessment in the major. In D. Heiland & L. J. Rosenthal (Eds.), *Literary study, measurement, and the sublime: Disciplinary assessment* (pp. 259–276). New York: The Teagle Foundation.
- Suskie, L. (2009). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Swender, E., & Duncan, G. (1998). ACTFL performance guidelines for K-12 learners. *Foreign Language Annals*, 31, 479–491. doi:10.1111/j.1944-9720.1998.tb00593.x
- Walvoord, B. E. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Walvoord, B. E. (2011). How to construct a simple, sensible, useful departmental assessment process. In D. Heiland & L. J. Rosenthal (Eds.), *Literary study, measurement, and the sublime: Disciplinary assessment* (pp. 335–352). New York: The Teagle Foundation.
- WEAVE. (n.d). Retrieved on August 30, 2013, from <http://www.weaveengaged.com>