THE COMMENTARIES

Online Foreign Language Education: What Are the Proficiency Outcomes?

CHIN–HSI LIN, Michigan State University

MARK WARSCHAUER, University of California, Irvine

Over the last few years, there has been an accelerating expansion of online language courses offered both by educational institutions and by commercial organizations. Courses by the latter, those that are free, reach a huge number of learners around the world. *Duolingo* had 30 million users as of April 2014 (von Ahn, 2014) and *Livemocha*, which is now owned by *Rosetta Stone*, had 13 million in 2012.

Yet, with so many people studying languages online, what are they actually learning? In an era of school accountability, student assessment is an essential part of education. In theory, assessment not only helps students evaluate their progress, but also helps teachers plan instruction and choose instructional materials, regardless of delivery mode. Understanding proficiency outcomes is even more important in language learning than in other fields, as many jobs, such as K–12 foreign language teaching, require candidates to reach Advanced Low proficiency in the target language according to the American Council of Teachers of Foreign Languages (ACTFL) guidelines (Burke, 2013). In addition, there is natural skepticism toward online language teaching (see discussion in Blake et al., 2008) since most online programs provide insufficient spoken contact among learners, peers, and teachers to foster oral proficiency (Goodfellow, Manning, & Lamy, 1999). A better understanding of the proficiency outcomes of online language education is needed to help language educators make informed decisions about whether to embrace it, and toward what ends.

ONLINE LANGUAGE LEARNING IN HIGHER EDUCATION

Most studies on online language learning have been conducted in higher education settings. Much of this research examines beginning-and/or intermediate-level courses, and typically

reports that online language education is at least as effective as its offline counterpart (Blake et al., 2008; Despain, 2003; Isenberg, 2010; Ushida, 2005). Chenoweth and Murday (2003), for example, found that online students in beginning French outperformed offline students in terms of written production (as measured by accuracy, use of transitions, overall development, and syntactic complexity), while achieving comparable results in listening comprehension, reading comprehension, grammar, and oral production (as measured by different sections in the final exam). Similarly, Chenoweth, Ushida, and Murday (2006) reported that students in beginning and intermediate French and Spanish courses in hybrid format (in which a portion of students' time is spent studying online, and the remainder in face-to-face learning) showed comparable progress to their offline counterparts in listening comprehension, reading comprehension, oral production, written production, grammar knowledge, and vocabulary (as measured by different sections in the final exam).

In the previously mentioned studies, language skills were assessed using achievement exams rather than standardized proficiency tests, and one apparent limitation is the focus on what was taught rather than overall proficiency. However, studies that employed proficiency testing had broadly similar results. Blake et al. (2008) demonstrated that hybrid courses and fully online courses in beginning and intermediate Spanish yielded similar oral proficiency outcomes on the standardized Versant test as did face-to-face courses. Isenberg's (2010) quasi-experimental study compared online and offline beginning German classes with regard to their learning outcomes, which were assessed via a standardized achievement test WebCape (which consists of questions in vocabulary, grammar, and reading comprehension), translation recognition tasks, grammaticality judgment tasks, and Simulated Oral Proficiency Interviews (SOPI) rated using Payne and Whitney's (2002) 50-point scale. She found that online and offline learners performed equally well in all four of these measures, and that improvements in performance on the

translation-recognition and SOPI tasks were greater among online students, though these differences were not statistically significant.

In short, published studies consistently report the effectiveness of online higher education teaching in beginning and intermediate language courses. However, as with any area of technology-based learning, a *file drawer* problem may be in effect, with those studies finding no benefits for online learning less likely to be published. In addition, if these findings are credible, questions about whether or to what extent they hold true in advanced language courses, or in LCTL courses, remain unanswered.

K–12 ONLINE LANGUAGE LEARNING

Five years ago, Means et al. (2010) lamented the limited number of rigorous studies examining the effectiveness of online learning in K–12 contexts. Little has changed, since the situation remains unclear due to lack of data (Watson et al., 2014). In K–12 public education in the United States, not all online courses report student achievement, as the requirement to do so varies by institution type. Most school districts have implemented or are in the process of implementing some form of online learning, ranging from supplemental online courses to fully online schools. Multi-district fully online schools are approximately as accountable as traditional brick-and-mortar schools, but other types of institutions and initiatives, including single-district online programs, consortium online programs, and private/independent schools, are not necessarily required to report student achievement data, let alone proficiency benchmark data.

A meta-analysis of online learning in K–12 settings by Cavanaugh (2001) found that, among all subject areas, only online foreign language courses yielded negative effects. Those findings, however, may be obsolete due to dramatic changes in K–12 online learning over the past 15 years. In addition, the absence of clear references to the studies that Cavanaugh analyzed makes it impossible to ascertain what may have caused these negative effects.

For the past year and a half, the lead author has been conducting research on Michigan Virtual School (MVS), a statewide supplemental program in which students take individual online courses while also being enrolled in a physical or cyber school within the same state. Most of the foreign language courses offered at MVS are self-adaptive and have only one deadline: by the final exam. Theoretically, a teacher can grade assignments handed in from the first to the last week of the term on the very last day. Students can also complete a course in a very short period of time. Such a design is worrisome because it may shield teachers from knowledge of their students' progress and thereby preclude early intervention when there are problems. Without regular formative assessment, learning relies heavily—arguably, too heavily—on students' capacity for self-regulation. It is even more worrisome, given that self-regulation is impacted by learner motivation, that 62% of online enrollments in virtual schools fall into the category of credit recovery (Queen & Lewis, 2011), a category that frequently overlaps with poorly motivated students. A recent study of North Carolina Virtual Public School prompts a further concern: Oliver, Kellogg, and Patel (2012) report that students enrolled in foreign language courses at the school had significantly lower perceptions of their online course than students taking other subjects had of theirs, while just 19% of intermediate and advanced foreign language students perceived that they learned as much online as they did in offline courses.

ONLINE LANGUAGE LEARNING WEBSITES

The extent to which out-of-school self-initiated online language learning can improve language proficiency is still an under-researched area. Currently, *Livemocha* and *Duolingo* do not incorporate proficiency tests at the point of course completion. However, *Duolingo* does have an adaptive placement test to help determine the appropriate course for a learner to enroll in. A recent study by Vesselinov and Grego (2012) indicated that 34 hours spent learning Spanish on *Duolingo* yielded similar results (as assessed by WebCape) to taking the first semester of Spanish in college, whereas *Rosetta Stone* users had to study between 55 to 60 hours to achieve similar results (Vesselinov, 2009). These findings should be interpreted with caution, however, as they assess achievement, not proficiency, and may not be generalizable to students of other languages or at other levels.

ONLINE PROFICIENCY TESTING

Most studies on measuring language learning in online environments focus on achievement, not proficiency. One reason may be the scarcity of online language courses at the advanced level. The costs and organizational challenges of online proficiency testing may also contribute to the lack of proficiency testing in online language instruction environments, especially given the physical dispersal of enrolled students (Blake et al., 2008; Malone &

Montee, 2010). Proficiency tests that can be administered remotely (e.g., online or telephone) might help administrators and researchers tackle this task. Currently, several proficiency tests are available for placement or research purposes through online or telephone administration, such as ACTFL's Assessment of Performance toward Proficiency in Languages (AAPPL) and Versant.

AAPPL is a web-based assessment of proficiency in interpretive listening, interpretive speaking, interpersonal listening and speaking, and presentational writing. The test takes about 2 hours to complete and is administered via a virtual instructor.

Versant tests, developed by Pearson Education, are available in six languages, including English, Arabic, Chinese, Dutch, French, and Spanish, and have been employed in several studies (Blake et al., 2008; Nielson, 2011). They can be administered over the phone or on a computer with Internet access and take 15 minutes to complete. Versant provides scores in overall proficiency, sentence mastery, vocabulary, fluency, and pronunciation. Versant tests evaluate the examinees' level of automaticity with the language (see Blake et al., 2008), which is different from the ACTFL OPI that assesses three modes of communication. Versant has a high test–retest reliability and strong correlation with the Oral Proficiency Interviews conducted by ACTFL and the Interagency Language Roundtable (Fox & Fraser, 2009).

WebCape, an online adaptive placement examination designed by Brigham Young University, is widely used in higher education. It is available in Spanish, French, German, Russian, ESL, Chinese, and Italian. WebCAPE has a high validity in Spanish (.91), German (.89), and French (.80), as well as high test–retest reliability in the same three languages (.86, .80, and .76, respectively) (see WebCape, n.d., for details). WebCape claims to have calibrated the test in accordance with ACTFL proficiency guidelines, but no technical report can be found on the website.

The Language Acquisition Resource Center at San Diego State University has developed an online oral proficiency test called the *Computer Assisted Screening Tool* (*CAST*; Language Acquisition Resource Center, n.d.) that claims to provide similar results as the ACTFL OPI. The Center for Language Education and Research at Michigan State University is currently developing a fully online oral proficiency testing program for LCTLs, which is expected to be released later in 2015 (Center for Language Education and Research,

n.d.). The Center for Applied Linguistics has developed a computerized OPI, which has comparable scores with SOPI (Kenyon & Malabonga, 2001; also see discussion in Kenyon, Malabonga, & Carpenter, 2001), but it has to be administered on a dedicated computer with its software license.

In summary, a number of different institutions or organizations are developing online or computer-based proficiency tests. Further development and evaluation of these tests will surely gain in the attention of applied linguistics in the coming years. This will be important for a number of domains, including teacher education, as we have not yet seen any studies indicating that online language courses produce the high proficiency outcomes required by K–12 foreign language teacher candidates.

FUTURE RESEARCH

Future research on online proficiency testing will thus be an important topic for applied linguistics in coming years. Documenting proficiency attainment in online courses in K–12 and postsecondary contexts will help to determine if students develop general communicative competence beyond the specific goals set up in the course. Research that deploys widely used extant proficiency measures, such as the ACTFL OPI and other validated proficiency tests, can be especially valuable.

A closer linkage is also needed between computer-based assessment and instruction (see discussion in Laurier, 2000). Accordingly, another aim of future research might be to examine ways of integrating proficiency testing into instructional and learning processes. One way this can be accomplished is by using online adaptive criterion-referenced proficiency tests, which evaluate test-takers' capability to perform a certain language task (Brown, 2005). *Duolingo*, for example, has an online placement test to determine users' level of proficiency in the target language and recommend courses appropriate to their level.

Another promising approach is *dynamic assessment*, employed by Oskoz (2005) to evaluate developmental outcomes that occurred through computer-mediated communication. Rooted in Vygotsky's (1978) concept of the zone of proximal development, dynamic assessment is a special approach that involves a teacher's intervention. Using a pretest–intervention–posttest paradigm, assessors examine the degree to which learners are able to acquire new skills and predict individuals' learning ability. One aim of future research might be to use computer-mediated

dynamic assessment to examine online learning development.

In summary, there is much work to be done, both to better assess language learning and proficiency online, as well as to exploit these assessments for adapting, personalizing, and improving online language learning experiences and outcomes.

REFERENCES

Blake, R. J., Wilson, N. L., Cetto, M., & Pardo–Ballester, C. (2008). Measuring oral proficiency in distance, face-to-face, and blended classrooms. *Language Learning & Technology, 12*, 114–127.

Brown, J. D. (2005). *Testing in language programs.* New York: McGraw–Hill.

Burke, B. M. (2013). Looking into a crystal ball: Is requiring high-stakes language proficiency tests really going to improve world language education? *Modern Language Journal, 97*, 531–534.

Cavanaugh, C. (2001). The effectiveness of interactive distance education technologies in K–12 learning: A meta-analysis. *International Journal of Educational Telecommunications, 7*, 73–88.

Center for Language Education and Research (CLEAR). (n.d.). *Speaking proficiency assessment.* Accessed 6 February 2015 at http://clear.msu.edu/assessment/online-speaking-proficiency-assessments/

Chenoweth, N. A., & Murday, K. (2003). Measuring student learning in an online French course. *CALICO Journal, 20*, 284–314.

Chenoweth, N. A., Ushida, E., & Murday, K. (2006). Student learning in hybrid French and Spanish courses: An overview of language online. *CALICO Journal, 24*, 115–146.

Despain, J. S. (2003). Achievement and attrition rate differences between traditional and Internet-based beginning Spanish courses. *Foreign Language Annals, 36*, 243–257.

Fox, J., & Fraser, W. (2009). Test review: The Versant Spanish Test. *Language Testing, 26*, 313–322.

Goodfellow, R., Manning, P., & Lamy, M.–N. (1999). Building an online open and distance language learning environment. In R. Debski & M. Levy (Eds.), *WORLDCALL: Global perspectives on computer-assisted language learning* (pp. 43–56). Exton, PA: Swets & Zeitlinger.

Isenberg, N. A. (2010). *A comparative study of developmental outcomes in Web-based and classroom-based German language education at the post-secondary level: Vocabulary, grammar, language processing, and oral proficiency development.* (Unpublished doctoral dissertation). The Pennsylvania State University, State College, PA.

Kenyon, D. M., & Malabonga, V. (2001). Comparing examinee attitudes toward computer-assisted and other oral proficiency assessment. *Language Learning & Technology, 5*, 60–83.

Kenyon, D. M., Malabonga, V., & Carpenter, H. (2001). Response to the Norris commentary. *Language Learning & Technology, 5*, 106–108.

Language Acquisition Resource Center. (n.d.). *Computer Assisted Screening Tool (CAST).* Accessed 6 February 2015 at https://larc.sdsu.edu/cast/

Laurier, M. (2000). Can computerised testing be authentic? *ReCALL, 12*, 93–104.

Malone, M. E., & Montee, M. J. (2010). Oral proficiency assessment: Current approaches and applications for post-secondary foreign language programs. *Language and Linguistics Compass, 4*, 972–986.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis of online learning studies.* Washington, DC: U.S. Department of Education.

Nielson, K. B. (2011). Self-study with language learning software in the workplace: What happens? *Language Learning & Technology, 15*, 110–129.

Oliver, K., Kellogg, S., & Patel, R. (2012). An investigation into reported differences between online foreign language instruction and other subject areas in a virtual school. *CALICO Journal, 29*, 269–296.

Oskoz, A. (2005). Students' dynamic assessment via online chat. *CALICO Journal, 22*, 513–536.

Payne, J. S., & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal, 20*, 7–32.

Queen, B., & Lewis, L. (2011). *Distance education courses for public elementary and secondary school students: 2009–10. First Look. NCES 2012–008.* Washington, DC: National Center for Education Statistics.

Ushida, E. (2005). The role of students' attitudes and motivation in second language learning in online language courses. *CALICO Journal, 23*, 49–78.

Vesselinov, R. (2009). *Measuring the effectiveness of Rosetta Stone.* Accessed 16 January 2015 at https://www.binghamton.edu/clt/pdf-files/Rosetta%20Stone%20Case%20Study%20Spanish%20at%20Queens%20College%202009%20Study%20Summary.pdf

Vesselinov, R., & Grego, J. (2012). *Duolinguo effectiveness study.* Accessed 14 January 2015 at http://static.duolingo.com/s3/DuolingoReport_Final.pdf

von Ahn, L. (2014). *Duolingo: Duolingo turns two today!* Accessed 10 January 2015 at https://www.duolingo.com/comment/3412629

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Watson, J., Pape, L., Murin, A., Gemin, B., & Vashaw, L. (2014). *Keeping pace with K–12 digital learning: An annual review of policy and practice.* Accessed 5 November 2015 at http://www.kpk12.com/wp-content/uploads/EEG_KP2014-fnl-lr.pdf

WebCAPE. (n.d.). *Details.* Accessed 8 January 2015 at http://www.perpetualworks.com/webcape/details