Testing an online English course: lessons learned from an analysis of post-course proficiency change scores

Rebecca Y. Jee[1]

Abstract. Voxy, an English-language-learning company, has developed a custom, in-house proficiency exam, the Voxy Proficiency Assessment (VPA), which is given to all learners at the beginning and end of their courses. Using Multinomial Logistic Regression (MLR), the impact of covariates, such as total learning activities completed and total number of resource types, on the likelihood of Voxy users showing a gain, maintenance, or loss in proficiency after twelve weeks of software use, was explored. Relevance of these findings for the design of online language learning activities as well as implications for assessment-related research within the L2 online learning environment will be discussed.

Keywords: blended learning, distance learning, language proficiency, proficiency over time.

1. Introduction

A Department of Education meta-analysis (Means et al., 2010) on blended learning and significant empirical research on the impact of blended language learning on learner gains (Blake, Wilson, Cetto, & Pardo-Ballester, 2008; Jee & O'Connor, 2014; Miyazoe & Anderson, 2010) offer support for using a blended method as an alternative to traditional classrooms. However, and given that it is difficult to have distance learners take pre- and post-course assessments (Nielson, 2014), thus contributing to the dearth of research on proficiency improvements with distance language learning, many online language courses and technological tools for

1. Voxy, New York, NY, USA; rebecca@voxy.com

How to cite this article: Jee, R. Y. (2015). Testing an online English course: lessons learned from an analysis of post-course proficiency change scores. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 281-287). Dublin: Research-publishing.net. http://dx.doi.org/10.14705/rpnet.2015.000346

distance language learning remain untested (Blake et al., 2008). Since pre- and post-course assessment results would offer valuable insights into the efficacy of a course as well as the test-takers' true abilities and learning needs, an English-language-learning company analyzed the results of its proficiency assessment that is given to learners at the beginning of their courses and again three months later. The data collected from these assessments allow Voxy to evaluate how various covariates impact proficiency change scores and to determine what leads to the greatest gains for English-language learners in their use of the online course.

2. Method

2.1. Learners

Learners in the study included 309 university students and working professionals learning English as a foreign language. All took at least two versions of the VPA, the in-product tool used to measure global English proficiency, discussed in more detail below. The learners currently reside in various countries worldwide and are native speakers of a wide range of languages, but the majority are native speakers of Portuguese and Spanish. Proficiency levels varied across learners as well, with level 1 being the lowest and level 12 being the highest (the highest Voxy proficiency level is 15). All learners used Voxy in conjunction with English courses they attended at a physical location.

2.2. Tools

Voxy is an online English-language-learning product that offers synchronous and autonomous instruction via an integrated, multi-platform system. Learners are able to access Voxy on computers or on mobile devices. Voxy was designed within the Task-Based Language Teaching (TBLT) pedagogic framework (Long & Crooks, 1992) and developed using established principles of instructed Second Language Acquisition (SLA) (Gass & Selinker, 2001) and effective online language learning (Nielson & Gonzalez-Lloret, 2010). In conducting this exploratory study, the VPA, Voxy resources, and VPA activities were focused on specifically.

2.2.1. VPA

Once learners provide information about what their interests and learning goals are, the next step is to take the VPA, a multiple-choice test that assesses grammatical knowledge, reading comprehension, and listening comprehension. Subsequent forms of the VPA are offered every three months, the results of which are used to

measure improvements in learner proficiency. All forms of the VPA are parallel so that the content and difficulty level of each of the items is the same across all forms. This way, test-takers' scores do not change over time based on difficulty level of the test item.

2.2.2. Voxy resources and activities

Each Voxy lesson is comprised of one resource and three subsequent activities. A resource could be any piece of authentic, real-world media like photos (Image resource), news articles (Article resource), and recorded conversations (Conversation resource) and videos (Video resource) of native English speakers completing everyday tasks. Activities target skills like vocabulary, reading comprehension, listening comprehension, pronunciation, and writing. It is important to note that depending on the first VPA score, learners will see more or fewer of certain resource types. For example, a more beginner-level learner will see more Image resources, which primarily focus on introducing various lexical items to learners, than a more advanced-level learner who will see more Video resources, which aim to familiarize learners with more authentic, real-life environments.

2.3. Procedures

Existing records on VPA answers, VPA scores, and Voxy usage for each learner were searched to identify learners who had taken at least two VPAs, since this is the only way to determine what kind of proficiency change took place. In order to examine the probability of category membership on more than two dependent variables in which each is nominal, a multinomial logistic regression (MLR) was used for this analysis. What types of Voxy resources would predict the likelihood of proficiency gain (moving up at least one Voxy level, and marked as "1") versus proficiency loss (moving down at least one Voxy level, and marked as "-1"), and the likelihood of proficiency maintenance (staying in the same Voxy level, and marked as "0") versus proficiency loss, were examined. The dependent variables were 1) gain, 2) maintenance, and 3) loss; the predictor variables were 1) total number of activities, 2) total number of image resources, 3) total number of article resources, 4) total number of conversation resources, and 5) total number of video resources. The multi-collinearity statistics showed that none of the predictor variables were highly correlated – if they had been, it would have been impossible to determine which of the variables was truly predicting proficiency change within the regression equation. Since the research is exploratory, the significance threshold was set at 0.1 rather than 0.05 (Cavana, Delahaye, & Sekaran, 2001; Merrill, 2013).

3. Findings and discussion

Figure 1 shows the actual breakdown of starting and ending Voxy proficiency levels for the 309 learners in this study. You can see that the majority of the learners are identified as Level 2. On the Voxy proficiency scale, this means that a learner is of a more beginner level. Table 1 shows the descriptive statistics for the total number of activities and resource types completed by all learners in this study. As aforementioned, the number of resource types varies greatly due to the nature of the breakdown of learner levels (i.e. many more Image resources than Video resources because there are more beginner-level learners than advanced learners).

Figure 1. Breakdown of learners' starting and ending proficiency levels

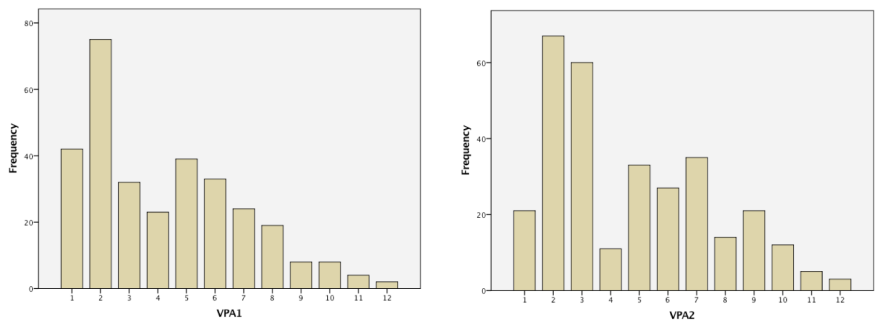

Table 1. Total number of activities and resource types completed

	N	Minimum	Maximum	Mean	Std. Deviation
Activities	309	0	3026	294.60	307.677
Image	309	0	2203	112.61	191.230
Article	309	0	1082	127.39	146.348
Conversation	309	0	197	34.02	33.642
Video	309	0	133	18.99	20.985
Valid N (listwise)	309				

Table 2 shows that for the grammar sub-skill, completing more activities lessened the likelihood of maintenance as opposed to loss, and all resource types increased the likelihood of maintenance. Completing more activities, that is to say, may not necessarily be the key to maintaining grammatical knowledge, but what may suggest maintenance is the type of resources learners complete. None of the predictor variables (i.e. total number of activities, image, article, conversation, and video resources) was significant with respect to the likelihood of gain. This

can most likely be attributed to the fact that Voxy lessons, in accordance with TBLT principles, do not focus on providing explicit grammar instruction when not relevant or necessary.

Table 2. Grammar sub-skill results

Parameter Estimates

VPA G[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	90% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
0	Intercept	-.027	.307	.008	1	.929			
	Activities	-.242	.127	3.632	1	.057	.785	.637	.967
	Image	.240	.127	3.567	1	.059	1.271	1.031	1.566
	Article	.238	.127	3.503	1	.061	1.268	1.029	1.563
	Conversation	.254	.128	3.957	1	.047	1.289	1.045	1.590
	Video	.231	.129	3.207	1	.073	1.260	1.019	1.559
1	Intercept	.890	.199	20.087	1	.000			
	Activities	-.008	.036	.051	1	.822	.992	.936	1.052
	Image	.007	.036	.042	1	.838	1.007	.950	1.068
	Article	.008	.036	.055	1	.815	1.008	.951	1.069
	Conversation	.019	.036	.272	1	.602	1.019	.960	1.081
	Video	-.013	.039	.112	1	.738	.987	.926	1.052

a. The reference category is: -1.

Table 3 shows that for the reading sub-skill, none of the predictor variables were significant with respect to the likelihood of gain or maintenance compared to loss.

Table 3. Reading sub-skill results

Parameter Estimates

VPA R[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	90% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
0	Intercept	2.000	.406	24.323	1	.000			
	Activities	.008	.078	.010	1	.921	1.008	.886	1.147
	Image	-.010	.079	.017	1	.896	.990	.869	1.127
	Article	-.003	.079	.002	1	.969	.997	.876	1.135
	Conversation	.006	.079	.005	1	.944	1.006	.884	1.144
	Video	-.016	.085	.034	1	.853	.984	.855	1.133
1	Intercept	1.336	.417	10.241	1	.001			
	Activities	-.047	.086	.302	1	.583	.954	.828	1.099
	Image	.049	.087	.315	1	.575	1.050	.910	1.210
	Article	.052	.087	.355	1	.551	1.053	.913	1.214
	Conversation	.051	.086	.345	1	.557	1.052	.913	1.212
	Video	.043	.093	.215	1	.643	1.044	.896	1.217

a. The reference category is: -1.

Table 4 shows that for the listening sub-skill, completing more Video resources decreased the likelihood of maintenance compared to loss, meaning that completing more Video resources may not necessarily be the key to maintaining listening comprehension skills. This may seem counter intuitive, given that Video resources should foster listening comprehension skills, since it is crucial for learners to understand the spoken content in these resources to complete the lesson. However, because many of the learners in this study were at proficiency levels that would

not enable them to see Video resources in the first place, this finding should be interpreted with caution – it would be more difficult to see proficiency gain or maintenance for this sub-skill if sufficient input was not provided in the first place.

Table 4. Listening sub-skills results

Parameter Estimates

VPA L[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	90% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
0	Intercept	.003	.245	.000	1	.990			
	Activities	.067	.051	1.715	1	.190	1.069	.983	1.163
	Image	-.067	.051	1.688	1	.194	.936	.860	1.018
	Article	-.069	.051	1.820	1	.177	.933	.857	1.015
	Conversation	-.053	.052	1.057	1	.304	.948	.871	1.032
	Video	-.092	.055	2.790	1	.095	.912	.833	.999
1	Intercept	.372	.208	3.217	1	.073			
	Activities	.011	.047	.057	1	.811	1.011	.935	1.094
	Image	-.009	.047	.038	1	.844	.991	.916	1.071
	Article	-.013	.048	.072	1	.788	.987	.913	1.068
	Conversation	-.014	.048	.087	1	.768	.986	.911	1.067
	Video	-.001	.050	.001	1	.977	.999	.919	1.085

a. The reference category is: -1.

4. Conclusion

The purpose of this exploratory study was to examine whether certain Voxy resource types or the extent of Voxy usage, as represented by the number of completed activities, would be able to predict a learner's proficiency gain, maintenance, or loss. While finding what would likely lead to proficiency gain was difficult, discovering which resource types would likely lead to maintenance and loss is sufficient in determining what improvements should be made to Voxy lessons to foster proficiency gain. It is also necessary to remember that the VPA is a *proficiency* assessment aiming to assess global proficiency and not an *achievement* test that assesses the learners' knowledge of what was taught in the lessons. Therefore, it would be difficult to determine exactly what learner behaviors when interacting with the Voxy product would most likely lead to proficiency gains. Due to the exploratory nature of the study, no additional information was collected on what learner activity takes place outside of the Voxy product or what happens during their English courses that take place in a physical location.

For future studies, perhaps knowledge of participation, or the lack thereof, in other learner activity, such as watching English language television shows and movies or having a native English-speaking pen pal, would contribute to a more complete picture of global proficiency change for these learners. Overall, the initial findings of this exploratory study show that it is difficult to predict what leads to proficiency

gain, but these findings are valuable in determining what steps can be taken to further improve the product.

5. Acknowledgements

I would like to thank Dr Katie Nielson and Mr Cesar Koirala for their support and guidance and Ms Megan Masters for setting the groundwork for this study. And it goes without saying that none of this would have been possible without the providence and support of Voxy.

References

Blake, R., Wilson, N., Cetto, M., & Pardo-Ballester, C. (2008). Measuring oral proficiency in distance, face-to-face, and blended classrooms. *Language Learning and Technology*, *12*(3), 114-127.

Cavana, R. Y., Delahaye, B. L., & Sekaran, Y. (2001). *Applied business research: qualitative and quantitative methods*. Australia: John Wiley & Sons Ltd.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: an introductory course*. London: Lawrence Earlbaum Associates.

Jee, R. Y., & O'Connor, G. (2014). Evaluating the impact of blended learning on performance and engagement of second language learners. *International Journal of Advanced Corporate Learning*, *7*(3), 12-16.

Long, M., & Crooks, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, *26*(1), 27-56. doi:10.2307/3587368

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: a meta-analysis and review of online learning studies*. US Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED505824.pdf

Merrill, R. M. (2013). *Fundamentals of epidemiology and biostatistics: combining the basics*. Burlington, MA: Jones & Barlett Learning.

Miyazoe, T., & Anderson, T. (2010). Learning outcomes and students' perceptions of online writing: simultaneous implementation of a forum, blog, and wiki in an EFL blended learning setting. *System*, *38*(2), 185-199. doi:10.1016/j.system.2010.03.006

Nielson, K. B. (2014). Evaluation of an online, task-based Chinese course. In M. Gonzalez-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: researching technology and tasks* (pp. 295-321). Amsterdam/Philadelphia: John Benjamins.

Nielson, K., & González-Lloret, M. (2010). Effective online foreign language courses: theoretical framework and practical applications. *The Eurocall Review*, *17*, 27-35.

Published by Research-publishing.net, not-for-profit association
Dublin, Ireland; info@research-publishing.net

Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy
Edited by Francesca Helm, Linda Bradley, Marta Guarda, and Sylvie Thouësny

Typeset by Research-publishing.net
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-28-5 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-29-2 (Ebook, PDF, colour)
ISBN13: 978-1-908416-30-8 (Ebook, EPUB, colour)

Legal deposit, Ireland: The National Library of Ireland, The Library of Trinity College, The Library of the University of Limerick, The Library of Dublin City University, The Library of NUI Cork, The Library of NUI Maynooth, The Library of University College Dublin, The Library of NUI Galway.

Legal deposit, United Kingdom: The British Library.
British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2015.