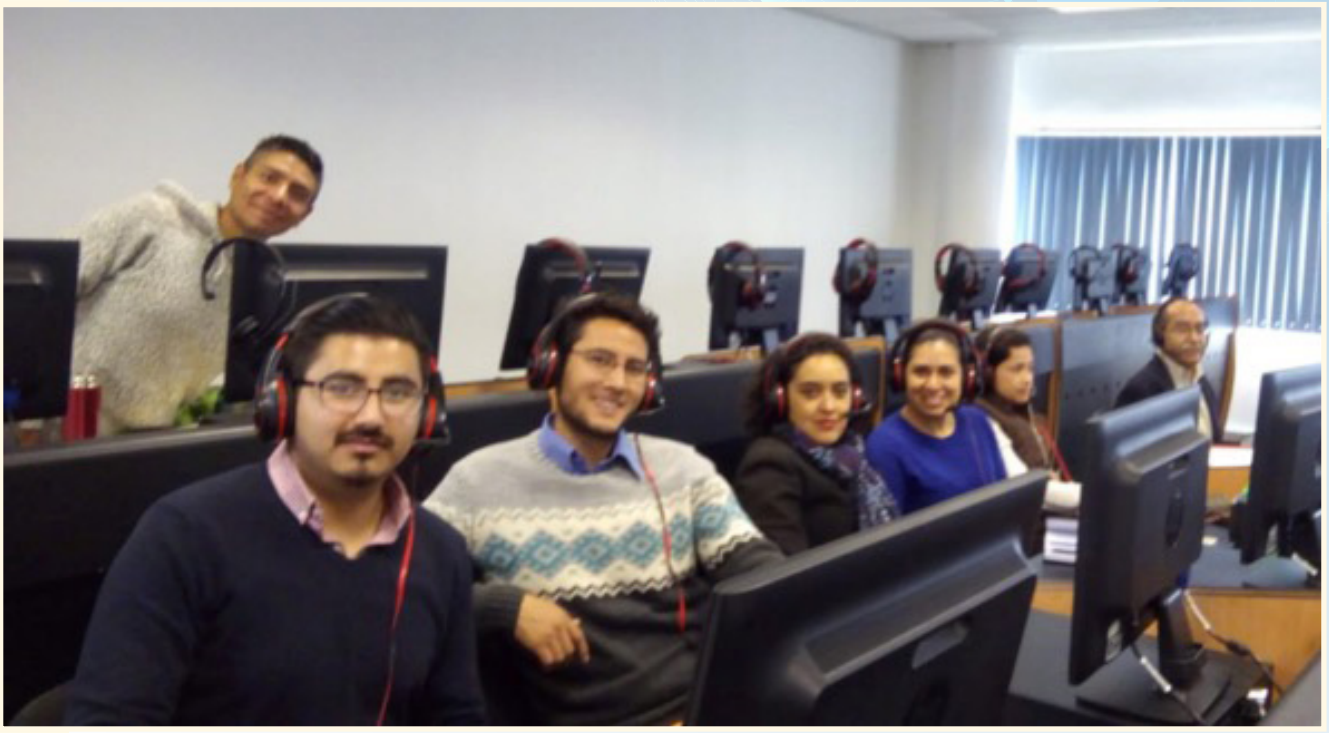
Case Study: UVM Program Using Alelo Enskill Increases Students' Spoken English Proficiency

Lewis Johnson, Ph.D.
Alelo Inc.
January 6, 2020



alelo®

Overview

Of the four language skills, speaking, listening, reading, and writing, speaking is by far the most important and the most difficult to develop (CUP, 2017). Many English language learners (ELLs) are reluctant to practice speaking, due to shyness or fear of making embarrassing mistakes. Spoken language activities in class are time-consuming and difficult for teachers to manage when many students are speaking at once.

Alelo Enskill®, an AI-driven learning platform, addresses these problems. Learners practice spoken language skills in AI simulations of realistic situations, populated by AI-driven avatars that speak and understand spoken language. Enskill helps students build self-confidence and make rapid progress toward proficiency. At the same time it reduces workload for teachers and makes more effective use of classroom and language lab time.

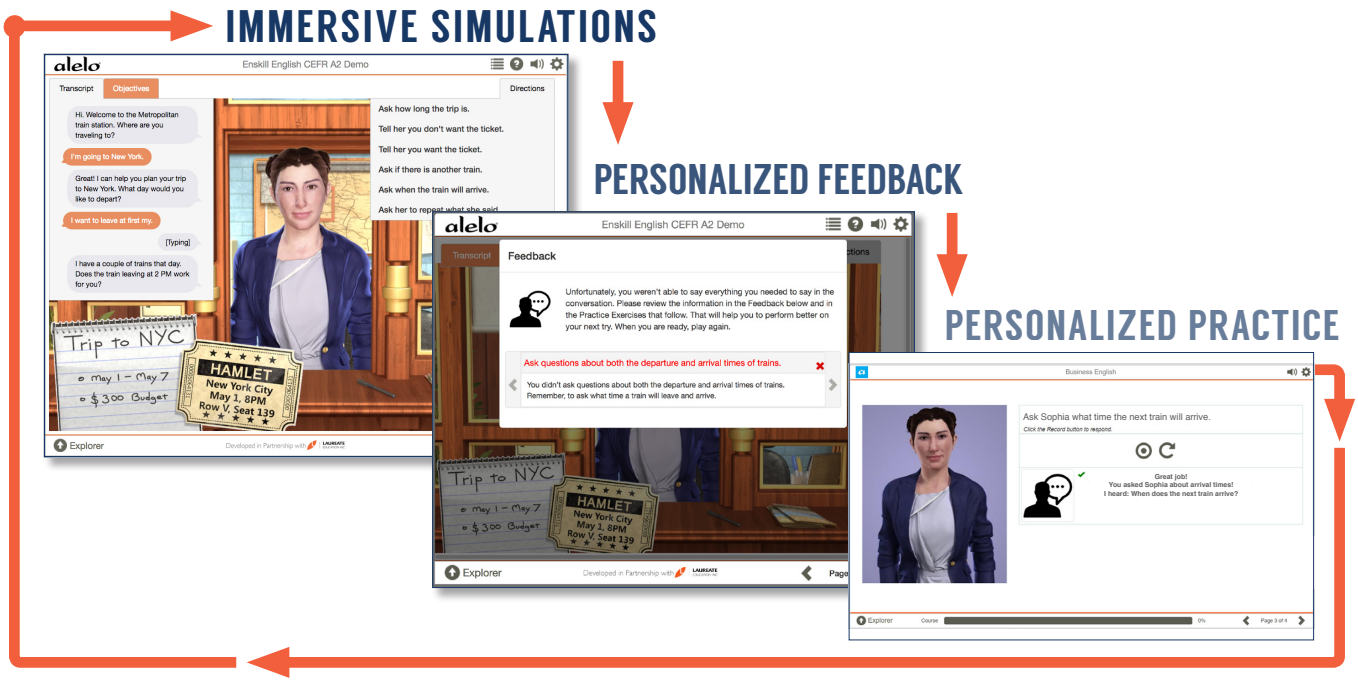
UVM (Universidad del Valle de México) Toluca Campus has integrated Alelo Enskill into its English curriculum. This report is an analysis of performance data collected by Enskill from learners participating in the UVM program. Enskill saves anonymized speech recordings and other learner data in the cloud, making archival analysis possible.

UVM learners practiced Enskill conversational simulations repeatedly, and their performance improved with practice. These findings are consistent with reports from the students themselves, who reported that Enskill helped them improve their spoken English skills. The analysis also showed that learners retained their spoken language skills well over time. Comparison of student performance before and after gaps in practice show that performance degrades very slowly, particularly for students at higher levels of language proficiency.

This study demonstrates the potential of cloud-based data-driven learning platforms to transform learning and assessment. Data collected by Alelo Enskill is used on an ongoing basis to monitor and assess learning, retrain AI models, and inform system development (Johnson, 2019). Real-time formative assessments make possible immediate feedback and personalized learning, and track each learner's progress toward proficiency.

Background: Enskill

Learners access Enskill through a web browser on their computer or mobile device. They converse with interactive characters by speaking into a microphone. The on-screen character interprets the learner's speech and responds, and at the same time evaluates the learner's communication skills. Thanks to Enskill's advanced natural language processing technology, learners can express themselves in a variety of ways and are not confined to a fixed script. At the end of each conversation Enskill provides feedback, including quantative metrics of performance, and recommends exercises for further practice.

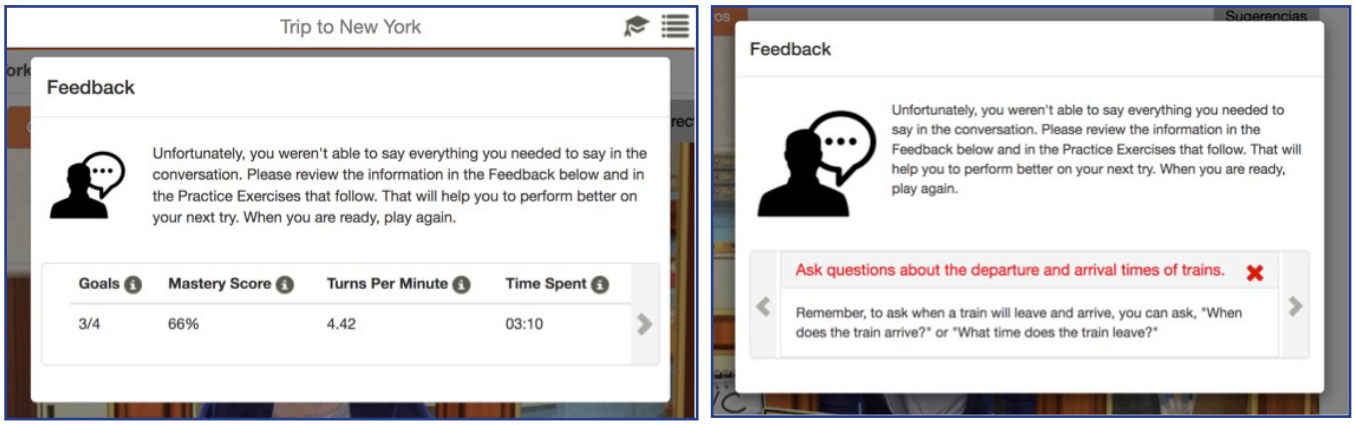


Practice and feedback cycle in Enskill

Enskill English courses are structured as a collection of task-based simulation modules. In each simulation the learner has a task to perform, such as buying a train ticket or getting directions to a destination. Simulations are organized into proficiency levels. The Common European Framework of Reference (CEFR) defines three levels for language proficiency: A (basic user), B (independent user) and C (proficient user). Each level is further subdivided as follows: A1, A2, B1, B2, C1, and C2. At the time of this writing Enskill English simulations are available for A1 and A2 levels, with more under development. Each level covers a semester-length course in English as a foreign language. There is a total of ten simulations at each level.

Each simulation is aligned with one or more can-do statements in the CEFR framework. For example the task of buying a train ticket is aligned with the can-do statement "Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time." Mastery of the task provides evidence that the learner can do what the can-do statement says. If a learner has mastered all of the tasks at a given level, it is likely that they have fully achieved that proficiency level.

In each task-based simulation the learner has a set of objectives to complete. Learners get feedback on their overall performance in the task, as well as on on each task objective. In the example below the debrief on the left shows that the learner achieved three out of four objectives; the objective that the learner failed to achieve is described on the right. Enskill tracks the overall time spent in the simulations, as a measure of learner effort, and the number of conversational turns per minute, as a measure of fluency. The turns per minute and the number of objectives completed are combined into an overall mastery score, which is normalized for each level so that learners at that level who have fully mastered the task should be able to perform at or near 100%.



Examples of simulation feedback

UVM Toluca Campus Blended Learning Program

UVM Toluca Campus has integrated Laureate Speaking Simulations into their English curriculum. Laureate Speaking Simulations are Enskill English courses developed for Laureate Education. At the beginning of each week the instructor goes over new language forms and structures in class. Afterwards they can practice these forms and structures in the Enskill simulations. They can practice on computers in the computer lab, and continue to practice at home as needed. At the end of each week learner participate in a speaking practice class, in which they apply the speaking skills that they have learned.

In initial trials UVM Toluca Campus divided students into sections, some of which used Enskill and some of which did not. After two months the teachers assessed the students' speaking skills using a rubric of speaking tasks. The teachers observed significant differences between the two groups. They have since adopted Enskill throughout their program.

Learners are encouraged to practice every simulation at least once. If they get a low mastery score they are encouraged to practice multiple times until they achieve a high mastery score.

Simulation Usage at UVM Toluca Campus

Data from 40 UVM users of the CEFR A1 simulations, and 67 users of the CEFR A2 simulations, in the period from August through December 2019, were selected for analysis. Of the 40 CEFR A1 users, 20 (50%) tried all 10 simulations at least once. Of the 67 CEFR A2 users, 37 (55%) tried all simulations at least once.

The following table shows the average usage at each level. On average learners practiced most of the simulations at least once, and practiced the simulations multiple times. Just 1 user (3%) at the A1 level practiced the simulations just once, and 8 (12%) at the A2 level practiced the simulations just once.

Level	Sims Tried (Avg)	Sim Runs (Avg)	Runs per Sim (Avg)	Max Score (Avg)
CEFR A1	7.10	22.08	2.96	62.28%
CEFR A2	7.31	17.72	2.27	75.69%

The average maximum mastery score that the A1 students were able to achieve over all simulations was 62.28%. For the A2 students the average maximum was 75.69%.

The data set included some high performers, as well as some low performers. On the high side, 4 A1 users (10%) achieved a mastery score of over 100% at least once and 10 A2 users (15%) achieved a mastery score of over 100% at least once. No student was consistently able to score over 100%, which indicates that they were all operating within the expected range for their level.

On the low side 8 users (20%) at the A1 level and 8 users (12%) at the A2 level never got a mastery score above 0%. These 0% scores are one reason why the average maximum scores were somewhat below 100%. 0% mastery scores result when learners fail to achieve any task objectives, and can occur for at least two reasons. Some had difficulty with the simulations and gave up without finishing. Others skipped through the simulations without saying anything. Enskill offers users the option of selecting utterances from a menu instead of speaking them; when they do this the simulation progresses but they get no credit for it in their mastery scores.

Performance Improvement and Decay over Time

To determine whether student performance improves with practice, we looked at each case where a user practiced a simulation more than once. We took the set of mastery scores for each simulation for each user, calculated linear regressions from the scores, and averaged the slopes of the linear regressions across the simulations. On average the A1 users who practiced simulations multiple times achieved an **increase of 10.97% in mastery score per trial.** The A2 users who practiced simulations multiple times achieved **an increase of 14.15% in mastery score per trial.**

We also investigated whether mastery scores decay after periods of inactivity. Within the study set there were 40 A1 users and 22 A2 users who practiced one or more simulations and then tried them again after a week or more of inactivity. In some cases the period of inactivity was over a month. The following table shows mastery scores before and after inactivity gaps between one week and one month in duration. In the period after the inactivity gap we examined the first non-zero score after the gap, as well as the average of all scores after the gap. For the A2 users there was

no decay in performance; the mastery scores after the inactivity gap were actually somewhat higher than they were before the inactivity gap. For the A1 users there was a slight performance decay. The difference may have been because the language skills of the A1 users were less well developed than that of the A2 users, and therefore were more vulnerable to decay.

Level	Score (Avg)	Avg. Before Gap	First After 1-Week Gap	Avg. After 1-Week Gap
A1	32.67%	31.39%	26.29%	29.47%
A2	44.16%	41.47%	45.03%	45.28%

The next table shows skill decay after an inactivity gap of one month or more. After this extended period both the A1 group and the A2 group showed some signs of skill decay.

Level	First After 1-Month Gap	Avg. After 1-Month Gap
A1	22.78%	24.55%
A2	23.43%	23.43%

The teacher in the Swedish trial also noted improvements in self-confidence, although it was hard to tell to what extent this was attributable to Enskill English and to what extent it was attributable to other learning activities that the students were engaged in.

Discussion

These analyses confirm that the UVM students were making effective use of the Enskill English simulations. Most students practiced the simulations multiple times, and the ones that did so improved their mastery scores.
After periods of inactivity performance decayed slowly, depending upon the level of proficiency attained. These findings are roughly consistent with second language attrition research (Weltens & Cohen, 1989), which suggests that as language learners achieve higher levels of proficiency their knowledge is less susceptible to attrition.
The current mastery score metric is not very sensitive at the low end of the scale, resulting in large numbers of 0% mastery scores. We therefore are adjusting the metric so that when learners engage in conversation even a little bit that is reflected in their scores. We are also evaluating the possibility of disabling the menu selection option, so that learners must always speak in order to make progress through a simulation.

Overall Results

UVM Toluca Campus teachers report that because of their practice with Enskill English, students come to speaking class better prepared and ready to engage in speaking activities. As the director of the English language program at UVM Toluca Campus, Erendira Yadira Carrera García, explains, "They feel more comfortable looking at an avatar or a picture that is not a real human being. They feel under pressure when they have the teacher there paying attention to every single detail, and sometimes we as teachers correct immediately." Daniel Vargas Iniestra, one of the instructors, commented: "It helps me improve my classes and it also makes my classes very very short and very very communicative."

Because students receive personalized instruction, they come to class better prepared. Ms. Carrera comments, "They have extra practice according to their needs. I think that is why it was so helpful, because it is focused on every student's needs."

Students also report that Enskill English helps them improve their English. Testimonials by UVM students, in English, may be found at the following Web address: https://youtu.be/mZvOrqJ3e9A. Here are some quotes from the students:

"Alelo helped me to improve my English speaking. So I recommend to use Alelo."

"This can help you a lot to practice, because you can repeat it as many times as are necessary."

"Alelo is a great tool for people who are shy or don't like to speak English in front of a group."

"You can practice a lot and improve a lot."

Conclusions and Recommendations

This analysis documents the benefits that UVM students are gaining from practice with Enskill English. Students are practicing the simulations, and the more they practice the more their performance improves.

The study illustrates the benefits of Alelo's data-driven development (D^3) process for developing learning technology (Johnson, 2019). Continuous access to anonymized learner data makes possible continuous learner assessment and system improvement.

Not all institutions using Enskill Enskill are experiencing such good results, particularly those where use of Enskill is optional. Other institutions should learn from UVM's success.

We are now extending Enskill's assessment capabilities and aligning them with standardized assessments such as TOEFL. That way students and teachers can track their progress toward proficiency, and also get specific feedback about where they can improve.

References

CUP (Cambridge University Press), 2017. Global Teaching Speaking Survey. Retrieved January 6, 2020 from http://www.cambridge.org/elt/blog/2017/11/06/teaching-speaking-survey-results/ .

Johnson, W.L. (2019). Data-driven development and evaluation of Enskill English. *Int. Journal of Artificial Intelligence in Education* 29(3), 425-457.

Weltens, B & Cohen, A.D. (1989). Language attrition research: An introduction. *Studies in Second Language Acquisition* 11(2), 127-133.