Digital Initiatives
Office of Innovation and Technology
Summer 2019

Stanford
GRADUATE SCHOOL OF
EDUCATION

Guadalupe Valdes Integrates Conversational AI into Online Course

Custom course and chatbot built by GSE IT trains new language coaches

To be understood is to belong. A man tries to connect with his co-workers, but is uneasy about using "dude" or "man" for fear of offending. He stands apart. A mother can't adequately communicate her epileptic son's condition to a 911 dispatcher. She waits, alone.

These are real cases, as documented by Guadalupe Valdes, Bonnie Katz Tenenbaum Professor of Education. Throughout her 28 years at Stanford, Dr. Valdes has endeavored to connect English-language learners (ELL) with meaningful and accessible language resources. At several points, she advised *Habla en el dia, habla en la noche* (*Habla*, for short), a program that pairs ELL Stanford employees -- such as groundskeepers, cafeteria workers, and custodial staff -- with Stanford undergraduates. The pairs regularly meet and conduct one-on-one conversations tailored to the linguistic and cultural needs of the adult learner.

By 2017, *Habla* had become an established service organization on campus. However, program leaders agreed that the learning experience needed iteration. "The overall sense was 'what we're doing is not good enough,'" notes Dr. Valdes. "The material needed work." Core members took a full year to identify improvements in materials and design, targeting more useful toolkits and strategies. More emphasis would be placed on relevance; there needed to be a stronger connection between content and each learner's background. And, crucially, the experience needed to be curated and preserved in a way that could weather the year-to-year turnover of a volunteer undergraduate organization. In Dr. Valdes' experience, "wisdom is finite since students graduate."

The solution: a highly interactive online course to train and track language coaches. Shawn Kim, Director of Digital Learning Solutions at GSE IT, assembled a team to produce a series of innovative and robust online modules. Katie Kormanik, an instructional designer, built out content based upon Dr. Valdes' vision and goals. Josh Weiss, Education Technology Specialist at GSE IT, researched web-based tools to accelerate learning. The project would be funded by Technology for Equity in Learning Opportunities (TELOS), a GSE initiative to financially support innovative ideas.

Dr. Guadalupe Valdes

The course, *Inglés Personal: Coaching Everyday Community English*, was designed over the fall 2018 term. The philosophy of the course was rooted in Dr. Valdes' background in second language acquisition. Conceptualizing language as a social process rather than a structure or form, all activities would be geared towards the background and needs of the learner. Determining the level of the student's current proficiency level would be key. "I can't support your language development if I don't know where you are," she points out. Empathy and learner-centered activity would be guiding principles of course development.

Designing a new experience

To provide more meaningful training, the team investigated web-based technologies that would make course content more interactive. At Shawn's suggestion, a chatbot, or AI-driven conversation interface, was considered. Chatbots like Siri and Alexa are coded to imitate human speech and respond based upon typical conversation patterns. Shawn and Josh

Stanford volunteers and adult learners pair up to improve English proficiency. The program recently underwent significant changes to better train language coaches.

hypothesized that if Siri and Alexa could be coded to converse at native English proficiency, perhaps a custom chatbot could also be coded to converse like a non-native speaker. Indeed, chatbots could provide new coaches a chance to interact with many types of English learners and to listen for characteristics that reveal proficiency levels. A simulated conversation could also function as a practice ground for determining the appropriate starting point for instruction, a key skill for language coaches.

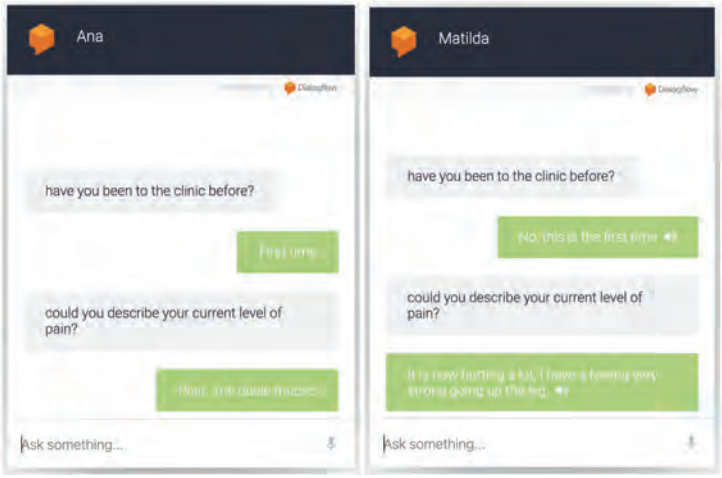
The chatbot design process then turned to purpose. Most chatbots are informational or transactional, and commonly utilized as answer banks to closed questions such as “How do I change my flight?” However, the goal of the *Inglés Personal* chatbot would be uniquely experiential and educational. In line with course goals, the purpose would be to provide a simulation for new language coaches to practice diagnosing English proficiency levels of ELL adults using a role-play. Rather than closed questions, the chatbot would respond to open and varied questions such as “Do you feel comfortable?” or “Can you tell me about your family?” The diagnostic framework on which coaches were trained emphasized context over grammatical correctness. Additionally, proficiency levels would reflect the chatbot’s ability to facilitate conversational flow, to use advanced strategies like circumlocution, and to deal with idiosyncratic speech. Finally, the rubric would entail three proficiency levels -- Beginners, Still Growing, and Fluent Functional. Accordingly, three chatbots would be developed to simulate each of the three levels.

	Beginners	Still growing	Fluent Functional
Understanding	Can understand very little. Struggles to understand personal information questions.	Can answer all personal information questions with ease. Can understand brief presentations on familiar topics when told what a presentation will focus on.	Can respond to general overview questions in English and offer details.
Speaking	Can use a few memorized phrases (address, age) in responding to personal information questions. Cannot role-play at any level and/or is hesitant to attempt to role-play.	Can role-play by circumlocuting and gesturing. Can follow some unexpected switches in role plays. Can ask for repetition etc. if needed.	Can converse in English with coach. Can make evident to coach what s/he can do with English in everyday life and in the workplace. Can communicate exactly what s/he needs to do in English.

The three chatbots were modeled on a language proficiency rubric designed by Dr. Valdes.

Making a learner-centered chatbot

Coding and technical development occurred in two stages. First, Josh and Dr. Valdes aimed to develop a rudimentary chatbot. After researching available technologies, Josh selected DialogFlow, a chatbot platform, based upon its machine learning features, maturity, and ease-of-use. At the same time, Dr. Valdes provided Josh with a rubric for characteristics of the three proficiency levels using a role-play based on a medical intake interaction. Based upon the rubric and some rudimentary scripts, Josh coded responses for each of the three proficiency levels into each of three discrete chatbot agents.



Each chatbot simulates a different level of English proficiency. Language coaches role-play with each chatbot to practice their evaluation skills.

Upon seeing the prototype of the chatbot, Dr. Valdes noted both the promise and the challenge of such a learning tool. On one hand, feasibility was a pleasant surprise. “It could really work -- we could actually get this close enough to be very useful to students.” On the other hand, authenticity would be a challenge, as open-ended questions required an extensive amount of scripted responses. For example, common expressions like “How are you feeling?” could be phrased by the language coach in a multitude of ways -- “What’s up?”, “You alright?”, etc. -- and each variation had to be coded in order for the chatbot to respond accurately. If this variation was not coded for, an inauthentic fallback response would occur (“I don’t know”), skewing the diagnosis.

As such, the second stage of development involved user testing and iteration towards authenticity. Josh and Dr. Valdes worked on a more robust rubric that drilled down into how each proficiency level would respond to types of questions, with an eye to idiosyncratic expression. Josh gathered more data to train the adaptive machine learning algorithm to recognize a broader variety of speech patterns and question types. Finally, Josh performed extensive user testing in order to better predict how new users would approach the interface.

The result was a narrower but deeper chatbot experience. Users were constrained as to how broad topics could range, but the depth, variation, and branching of conversation topics improved markedly. As a final push towards authenticity, Wilson Wang of GSE IT also developed custom code outside of the platform in order to enhance the chatbot's capacity at circumlocution and conversational flow.

The final step was implementation. As instructional designer, Katie embedded the chatbot directly into the Edx course site so that users could access instructional materials and the chatbot side-by-side. The course was launched April 1, and is currently underway, with a fresh cohort working through the course and preparing for their roles as language coaches.

Lessons learned

When looking back on the experience, Dr. Valdes points to some crucial takeaways.

Establish priorities: What are you driving towards? Content knowledge? Interactivity? Authenticity? Empathy? Identify essential cognitive or social objectives ahead of time. This will focus the project's energy on the most impactful segments along the learner experience according to your vision.

Acknowledge challenge alongside promise: New technologies can be alluring, but require familiarity and technical expertise to employ effectively. Sometimes the technology is just not mature enough to accomplish your entire vision. Prototyping and user testing can bring these strengths and weakness into focus more rapidly.

For more information about chatbots, course design, or emerging education technologies, reach out to GSE IT Digital Learning Solutions at instructionalsupport@stanford.edu.

Data Science Invigorates New Areas of Education Research

The recent profusion of data and quantitative analysis tools sets the stage for deep understandings into how and why learning happens

When interpreting large-scale behavior, Dr. Benjamin Domingue, Assistant Professor at Stanford Graduate School of Education, looks to data science for insights. Armed with data sets and a laptop, he is able to run analyses and statistical experiments from his desk, tweaking ideas and theories that formerly would have required teams of researchers and programmers. "Knowing how to work with the data—having some sense for what information can and can't be conveyed—is powerful," he emphasizes.

Such is the basis for data science, an emerging field that melds statistics, computational power, and domain expertise. Once a stronghold of computer scientists and mathematicians, data science is increasingly common in finance, medicine, geology, and even sports, as the movie *Moneyball* attests.

Education research can be added to this list. As more classroom interactions are performed digitally, more potential data becomes available. This can be as simple as counting how many times a student logs into a website or as complex as analyzing socio-emotional traits based upon patterns in a MOOC discussion forum. This trend coincides with the rise of so-called Big Data, set off by recent improvements in not only data volume and variety, but also networks and analysis tools that connect and sort information.

One example is Wildflower Montessori Schools. Its founder, Sep Kamvar, endeavored to create a modular and responsive learning environment for preschool children. To this end, he provided students with sensor-embedded shoes that tracked how each student moved and interacted with their natural environment. After the data was collected and analyzed, he discovered patterns in how students played and learned with objects, instructors, and each other. This data, combined with traditional Montessori observation techniques, informed a rich perspective into each student's tendencies in the classroom.

Dr. Benjamin Domingue

Advanced data science techniques can further derive insights from school-based relationships. Dr. Domingue leverages the R programming language to understand connections among genomes, social networks, educational and occupational attainment, psychological characteristics, and more. In a recent study, he investigated genetic similarity in school-based friendships, finding "that macro-level forces, such as school assignment, are a prime source of genetic similarity between friends". Furthermore, the genes of one's social peers are predictive of how far one goes in school; similar findings are also touched on in another recent study. In contrast, the genes of one's peers are largely irrelevant for physical attributes like height and BMI.

One key to research was the procurement and management of unique data sets. "The study utilized a unique data resource, the National Longitudinal Study of Adolescent to Adult Health, that contains data from thousands of respondents originally in school in the mid-90s," explains Dr. Domingue. "At baseline, school social networks were

At Wildflower Montessori School in Massachusetts, young learners' natural inclinations are revealed through instructor observation coupled with pattern recognition software.

ascertained; in the course of follow-up, respondent biospecimens were collected and subsequently genotyped." This merging of data resources provides a new opportunities to study how genetic effects may be socially contextualized.

Data science is increasingly being leveraged in online environments to inform and remix instruction. Commonly referred to as learning analytics, this branch of data science focuses on patterns of cognition that can be captured on a digital interface. Learning analytics takes cues from theorists like Jean Piaget, a developmental psychologist, and B.F. Skinner, a behaviorist and social philosopher, in an effort to determine optimal learning conditions on a person-by-person scale.

Although a relatively new field, learning analytics has accelerated thanks to research and resources. Stanford's Lytics Lab has showcased projects such as *Epistemic cognition: A promising and necessary construct for enriching large-scale online learning analysis*, which recommends methods for modeling knowledge states based upon online interactions. In addition, cloud-based data processing tools like Tableau have made it much simpler to beautify data, facilitating data-intensive conversations among educators and researchers. And innovative coding languages like D3.js allow for interactive data visualizations that illuminate trends in customizable ways.

While powerful, data science is not a magic fix. It can take time to familiarize with the tools and underlying statistical frameworks. Dr. Domingue emphasizes the importance of process. "Learning to ask simple questions to which you can provide robust answers is really important," he notes. "I think of it like a sculpture. I first try to get the overall shape of the thing...I take the rock and turn it into the raw shape, and then I do the detail work." Over time, insights emerge as the combination of statistical heft, computational power, and domain knowledge cohere: "In that process, you are developing intuition and fluency for working with the data and also generating additional insights into what the data is trying to teach you."

Increasingly, there is demand for making sense of unstructured data. These data are not easily sorted or labelled, and often arrive as a mish-mash of words, numbers, and even pictures. Gartner, a research firm, estimates unstructured data makes up for 80% of all information in business enterprise systems. This number could reach even higher for education-facing systems, considering the qualitative and media-rich nature of online learning. As such, language analysis has emerged as a major focus, embodied on campus by groups like the Center for Spatial and Textual Analysis, which aims to "apply digital tools and methods to create new knowledge in interdisciplinary humanistic inquiry."

Some researchers are going one step further and making sense of data with the help of pattern-finding algorithms. Natural language processing (NLP), an offshoot of artificial intelligence, parses language for structure and meaning. Zach Pardos, Assistant Professor in the Graduate School of Education and School of Information at UC Berkeley, utilizes this capacity in order to interpret how learners misunderstand concepts. Dr. Pardos tuned an NLP system to ingest sequences of answers from a tutoring program and output clusterings of incorrect student responses. Over time, the NLP system can map out patterns, in this case the relationship between wrong answers.

Tools like Tableau and d3.js can visualize learner data via custom code.

NLP techniques could have broad applications for learners and instructors. The relationships involved in linguistics often parallel those in broader social contexts. "If you think of language as behavior, then a model of language may be an appropriate model for other manifestations of expression," notes Dr. Pardos. "Your course selection behavior is, in part, an expression. It's an expression of your conception of those things at that time." Moreover, in an expanding field such as data science, pedagogy may be constructive as a framing tool. "Education theory can provide that useful element of constraint to help guide the process of knowledge construction."

For instructors and researchers interested in working with data science, GSE IT offers consultation and advisory services. If you would like to explore tools, storage, visualizations, security, or any part of the data science workflow, please reach out to instructionalsupport@stanford.edu.

Digital Initiatives highlights innovative research, design, and implementation around learning technologies at Stanford Graduate School of Education.

Written by Josh Weiss (josh.weiss@stanford.edu) and published by The Office of Innovation and Technology. For more information, visit http://gse-it.stanford.edu

Stanford | GRADUATE SCHOOL OF EDUCATION