

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Computer Assisted Learning in a (Dis-)Connected Age : : Challenges and Approaches to Digital Education and Equal Access

Permalink

<https://escholarship.org/uc/item/3jh4h76w>

Author

Alexandre, Jamie D.

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Computer Assisted Learning in a (Dis-)Connected Age:

Challenges and Approaches to Digital Education and Equal Access

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Cognitive Science

by

Jamie D. Alexandre

Committee in charge:

Professor Jeffrey L. Elman, Chair
Professor Grant Goodall
Professor James D. Hollan
Professor David Kirsh
Professor Rachel Mayberry

2014

The Dissertation of Jamie D. Alexandre is approved, and it is acceptable in
quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2014

Dedication

For Isabelle

You found me in the darkness

And gave me the courage

To step into the light of day

To live the life I wanted to live

With you there by my side

Every step of the way

Table of Contents

Signature page.....	iii
Dedication.....	iv
Table of Contents.....	v
List of Figures.....	ix
Acknowledgments.....	xiv
Vita.....	xxi
Abstract of the Dissertation.....	xxii
Chapter 1: Introduction and overview.....	1
What is the purpose of education?.....	1
A brief history of educational inclusion.....	2
Global barriers to equal educational access.....	5
Overview of the questions and challenges addressed in this dissertation.....	6
Chapter 2: Tutoring, Computer Assisted Language Learning, and ESL Genie.....	9
Introduction.....	9
Human and computer tutoring.....	10
The power of a tutor.....	10
Equality of access to tutoring.....	11
Automated tutoring.....	12
Computer Assisted Language Learning.....	13
Order of acquisition effects.....	13
Order effects and pedagogy.....	15
Approaches to CALL.....	16

Item Response Theory and methods for adaptive testing.....	18
Overview.....	18
Structured knowledge domain.....	20
Inference in a Bayesian IRT system.....	22
ESL Genie: An investigation into adaptive error diagnosis in ESL learners.....	26
Introduction.....	26
Quiz interface design.....	30
Methods.....	34
Exclusion criteria.....	40
Demographics.....	41
Analysis.....	42
Results from Item Response Theory testing.....	45
Exploring information sources for estimating prior ability.....	54
Results from the self-paced reading task.....	58
ESL Genie conclusion.....	64
Chapter 3: Online education and the digital divide.....	66
Open Educational Resources.....	66
You say you want a revolution: Khan Academy, and the MOOCs.....	67
The online learning revolution.....	67
Pedagogical explorations: self-paced learning and flipping the classroom.....	68
The digital divide.....	69
The historical divide.....	69
The days of proto-Internet.....	70

The emergence of the Internet.....	72
Access to the Internet.....	73
Where is the “cloud”?.....	75
Acknowledging the digital divide.....	78
Chapter 4: Taking the online learning revolution offline with KA Lite.....	81
Identifying opportunities and approaches.....	81
The rise of low-cost computing.....	82
The parallel revolution.....	82
One Laptop Per Child.....	83
Android and the Aakash.....	84
The Raspberry Pi.....	87
Design considerations for KA Lite.....	88
Be able to install and work offline.....	88
Basic features like video watching and exercise interaction.....	89
Progress tracking, points, and coach reports.....	90
Support content updates.....	91
Data synchronization (eventual consistency).....	91
Cross-platform support (portability).....	93
Running on low-powered and inexpensive hardware.....	94
Ease of installation process.....	95
Internationalization.....	95
KA Lite deployment contexts.....	96
Classrooms: Burmese monastery school.....	96

Classrooms: Nalanda Project.....	97
Orphanage.....	101
Correctional facilities.....	102
Empowering local community-driven content creation.....	107
Education and culture.....	107
Linguistic diversity of Open Educational Resources.....	109
Looking forwards.....	110
Chapter 5: Whither now?.....	112
ESL Genie.....	112
KA Lite.....	114
Appendices.....	117
Appendix A: Tagging scheme for annotating errors in essays.....	117
Appendix B: 84 known countries where KA Lite has been installed.....	118
Appendix C: Linguistic diversity and language proportions in Open Educational Resource repositories.....	119
References.....	121

List of Figures

Figure 1: Examples of successful and failed parses in a unification grammar....	17
Figure 2: Example 3PL (3 parameter logistic) sigmoidal item response function	19
Figure 3: Hypothetical graph structure used to predict response measures.....	21
Figure 4: Distributions and links between latent proficiency and response measures.....	22
Figure 5: Effect on the posterior proficiency estimate of answering questions with varying difficulty either correctly or incorrectly.....	25
Figure 6: Effect on mean estimated proficiency of question difficulty and discrimination (cooler colors mean a lower-variance, i.e. more confident, posterior).....	26
Figure 7: Estimated distribution of first languages in the AWPE essay corpus....	28
Figure 8: AWPE transcription interface with magnifying glass.....	29
Figure 9: Distribution of error tags in the transcribed essay corpus from ESL and non-ESL writers.....	30
Figure 10: Versions of color-coding correctness on questions, before user testing (top) and after user testing (bottom).....	33
Figure 11: Inline “fill in the blank” question display on hover.....	34
Figure 12: Subject flow through the stages of the experiment.....	35
Figure 13: English background information on the profile page.....	36
Figure 14: Language autocompletion tool to gather detailed information about languages spoken.....	37
Figure 15: Language proportion interface.....	38

Figure 16: Image prompt for the essay-writing component.....	39
Figure 17: Introduction to self-paced reading task.....	40
Figure 18: Examples of dummy questions, designed to detect subjects who were not attending to the task.....	41
Figure 19: Language backgrounds of included participants.....	42
Figure 20: Correctness as a function of response time.....	44
Figure 21: Item response function curves for all of the 308 included questions, each graph having an x domain of $(-4, 4)$ and y range of $(0, 1)$	46
Figure 22: Exemplar curves from questions in our test bank.....	47
Figure 23: Question with negative IRF slope, more often answered correctly by lower proficiency subjects.....	48
Figure 24: A non-diagnostic question, characterized by being answered incorrectly equally often by the top performers and the bottom of the barrel.....	48
Figure 25: Average error (across subjects) of leading question selection algorithms.....	51
Figure 26: Similarity between sequences generated by different question selection algorithms.....	53
Figure 27: IRT curve and question text for Q85, picked as the first question by both entropy minimization and information maximization.....	54
Figure 28: Relationship between self-rated proficiency and posterior estimate of actual ability.....	55

Figure 29: Effect of using self-reported proficiency as the prior mean for subject #93, who accurately provided a high self-assessment (self-rating of 0.826, posterior of 0.815).....	56
Figure 30: Effect of using self-reported proficiency as the prior mean for subject #71, who was overly modest in self-reporting ability (self-rating of -1.791, posterior of 1.409).....	57
Figure 31: Median self-paced timings (in ms) across the sentence with error at “realize”, binned by low and high proficiency.....	61
Figure 32: Median self-paced timings (in ms) across the sentence with error at “realize”, binned into four different groups by increasing proficiency.....	62
Figure 33: Scatterplots and correlations between self-paced reading syntactic error response times, and our posterior estimated ability (upper) and self-reported proficiencies (lower).....	64
Figure 34: Mainframe topologies.....	71
Figure 35: Network topologies in the continuum from (a) centralized networks, to (b) decentralized networks, to (c) distributed networks (from Baran, 1964).....	73
Figure 36: Global distribution of Internet users by country as a percentage of population.....	74
Figure 37: Scatterplot showing the relationship between student-teacher ratios and Internet connectivity in a country.....	75
Figure 38: Location of Google’s data centers.....	76
Figure 39: Geographic distribution of servers hosting the 100 most popular websites, when accessed from a computer in India.....	77

Figure 40: Internet penetration levels in low-income, middle-income, and high-income countries.....	80
Figure 41: Worldwide Smartphone Sales, from "Gartner Smart Phone Marketshare 2007-2013"	85
Figure 42: StatCounter Global Stats: Top Mobile & Tablet Operating Systems Per Country, Nov 2013.....	86
Figure 43: Raspberry Pi Model A board.....	88
Figure 44: Modes of delivering software/content to the final destination.....	89
Figure 45: KA Lite homepage, showing entry points into the content.....	90
Figure 46: KA Lite point system and coach reports.....	91
Figure 47: Interface for downloading videos and subtitles (left) and portion of video showing Chinese subtitles (right).....	91
Figure 48: Peer to peer synchronization network.....	92
Figure 49: Our simplification of the traditional web stack to make it into a pure Python codebase that can be installed and distributed easily.....	93
Figure 50: A hardware configuration using the Raspberry Pi as a local Wi-Fi access point, powered by a solar-charged cellphone charger battery.....	94
Figure 51: KA Lite main page translated into Bulgarian.....	95
Figure 52: Students in a Burmese monastery school using KA Lite.....	96
Figure 53: Hardware configuration used in Nalanda Project classrooms, with 1 Raspberry Pi wifi access point server running KA Lite, and 35 Aakash tablets, one for each student (5 shown).....	98

Figure 54: Students in an Akanksha school in India, using KA Lite on Aakash tablets.....	99
Figure 55: Children using KA Lite on donated laptops at an orphanage in Cameroon.....	101
Figure 56: Entrance to a prison we visited near Boise, Idaho.....	102
Figure 57: The sequential process of entering a secure area by means of a sally port. 1) Outer door is opened, and visitor passes into the sally port. 2) Outer door is locked. 3) Inner door is opened, and visitor exits from the sally port holding area into the secure area.....	106
Figure 58: Digital sally port, where server is alternately connected to the classroom (upper picture), and to the Internet (lower picture).....	107

Acknowledgments

By far the best part of my journey through graduate school, in the Department of Cognitive Science at UC San Diego, has been the people with whom I've had the chance to interact and work. The department itself served as a perfect incubator for the non-traditional projects I ended up pursuing, simultaneously instilling in me a scientific rigor and also fostering my exploratory and practical nature. I learned and benefited from countless people during my time here, only a few of whom I will have space to mention here.

Firstly, I would like to thank Jeff Elman, my adviser and mentor, who has been consistently generous with his time and avuncular advice, and ever supportive of the twists and turns of my academic meandering. Without Jeff's unwavering trust, and his support (both intellectual and financial), I would not have been able to follow the path I did and get to where I am today. Jeff continues to be a close mentor in my work with the Foundation for Learning Equality, as a founding board member and key adviser and advocate.

I am very grateful also to the rest of my dissertation committee -- Jim Hollan, David Kirsh, Grant Goodall, and Rachel Mayberry -- for being willing to let me pursue a risky and non-traditional path in my proposed work, and for their encouragement and advice along the way.

For training and mentorship, particularly through my first two years and my second year project, I am indebted to Marta Kutas, Tom Urbach, Lindsay Crissman, and the rest of the Kutaslab. Marta's intellectual rigour combined with

her intense caring for everyone around her has been a constant inspiration throughout my graduate career.

Beverley Walton, Thanh Maxwell, Mark Wallen, and the rest of the Cognitive Science support staff helped to make working here all the more enjoyable and productive, and I am indebted to all their help over the years I've spent in the department.

During my first experience mentoring research assistants, in my second and third years, I probably learned even more than my fantastic assistants Ryan Cordova, Wai Ho Chiu, and Frances Martin, who helped to run almost 100 subjects through a 2 hour experiment, and contributed to the analysis and interpretation of the results that formed the basis of my second year project. They were a joy to work with! My second year and third year projects also benefited greatly from mentorship and advice by Roger Levy of the Linguistics Department. Thanks also to fellow graduate students Leo Trottier and Nathaniel Smith for encouraging me to pick up Python again during this period, Micah Bregman for creating thisCourse with me to support the habit, and Jeremy Karnowski and Richard Tibbles for helping to run the Awesome Python Experience workshops. I also benefited greatly from participating in the Sequential Analysis Working Group with Kaya de Barbaro, Deborah Forster, Chris Johnson, Whitney Friedman, Paul Ruvolo, Paul Rodriguez, Nan Renner, Linda Phan, Gwen Littlewort and Adam Fouse.

I am indebted to the Center for Research in Language for employing me as Speaker Series Coordinator and Technical Report Editor for over 4 years. I

learned so much from being a part of the CRL community, and greatly enjoyed working with Robert Buffington, Margaret Paulson, Victor Ferreira, Marta Kutas, and so many others.

Many thanks to Seana Coulson and Andrea Chiba for their guidance and mentorship through our third year projects.

The ESL Genie project is famous for having converted my office from a lonesome graduate student cave into a hive of bustling undergraduate developers, designers, analysts, and linguists, from which it has happily never recovered. This amazing samosa-fueled team, which taught me so much about language, user experience design, and how to be part of something awesome, included: Wesley Hsu, Emilie Seubert, Irvin Fong, Clarice Robenalt, Guan Wang, Matthew O'Rourke, Vicky Tu, Cory Thorpe, Dylan Barth, Raelene Diega, and Stacey Sugiono.

I am so grateful to have had the opportunity to spend the summer of 2012 surrounded by the stupendously skilled, inspirational, and hilarious folks at Khan Academy. Thanks to Jessica Yuen, and conspirator Nicholas Butko, for encouraging me to get my foot in the door. Thanks to my official dev mentors -- Desmond Brand and Tom Yedwab -- as well as my unofficial dev mentors -- Ben Kamens, Ben Alpert, Ben Komalo, Ben Eater, Stephanie Chang, David Hu, and Marcia Lee -- for teaching me so much, and always being patient and eager to help. Thanks to Jascha Sohl-Dickstein for organizing a weekly journal club with consistently thought-provoking papers and debate, and to Elizabeth Slavitt for bringing us together for Big Ideas discussions that never failed to inspire. Thanks

to Esther Cho, without whom Khan Academy could not function, for making every day a joy and for introducing Sarah Singer into our lives. Thanks to my fellow Khan Academy interns, with whom I shared a fantastic and inspirational summer -- Ankit Ahuja, Ben Alpert, Leah Alpert, Drew Bent, David Hu, Jessica Liu, Omar Rizwan, Arun Saigal, Mary Tao, Dylan Vassallo, Mark Wittels, Jamie Wong, Josh Zimmerman -- and especially Emily Eisenberg, who was my SkunkWorks co-conspirator on the Khanberry Pi project that would later evolve into KA Lite. Thanks to everyone at Khan Academy -- especially Matt Wahl, Karl Wendt, Rishi Desai, Sundar Subbarayan, Kitt Hirasaki, and Bilal Musharraf -- for their excitement and encouragement in response to our Khanberry Pi demo, inspiring us to carry the work further. And many, many thanks to Bilal and to Maureen Suhendra for inviting the KA Lite team up to Khan Academy in February of 2013 to present our work, and for their continued advocacy, involvement, and support. Thanks to Craig Silverstein and James Irwin for their advice and help with internationalization of KA Lite, Marcos Ojeda for his adapted logo for KA Lite, and Minli Virdone (as well as Jessica, again) for advising on strategy and partnerships. And thanks to everyone else who was a part of making Summer 2012 one of the best summers of my life -- Jason Rosoff, Michael Burdick, Salman Khan, Shantanu Sinha, Beth Harris, Steven Zucker, Charlotte Koeniger, Chris Klaiber, Jace Kohlmeier, Joel Burget, John Resig, Vi Hart, and Yun-Fang Juan.

I feel immense gratitude each and every day for being able to be part of the group of people that is the Foundation for Learning Equality. Emerging out of

the cauldron of excitement and inspiration that our early release of KA Lite evoked, this team of amazing people has become my second (or, probably more accurately, my 7th) family, and consistently inspires and humbles me with their passion, dedication, wit, and caring. Dylan Barth, we could never have gotten this far without your dogged persistence, high standards, constant hard work, and your wisdom that goes way beyond your years; I have relied on you consistently throughout this epic journey, and you have never failed to deliver. Richard Tibbles, your willingness to put what you were doing to the side and devote time to KA Lite, back while it was still in its earliest stages, was very touching to me both as a vote of confidence and in terms of the huge material contributions you made and have continued to make; you are an intellectual and personal inspiration to me. Guan Wang, you never cease to amaze me with your nuanced analyses of human behavior and perception, and the intense thought that you put into your work and supporting those with whom you interact; you have been a huge part of both ESL Genie and now KA Lite, and we're lucky to have you with us. Ben Cipollini, when you do something, you go all in! I was so excited to have you jump on board, and your contributions to both the code and the organization's development have been of astronomical proportions; I'm so thankful for the big role you have played in helping us to get organized, to ask ourselves the hard questions, and to figure out how we're going to be able to stay around for the long-term. Elizabeth Vu, you have helped us to flourish externally -- leveraging your crystal clear communication and deep thinking -- while also keeping us connected and cohesive as a team internally, and I am so grateful to

be able to work with you. Rui Malheiro, from the moment you heard about KA Lite, all the way on the other side of the world, you have thrown yourself at it wholeheartedly despite the challenges you face in your own life; you are an inspiration to the rest of the team, and your contributions continue to help people everywhere. Aron Fyodor Asor, you joined us as a volunteer intern, flying down to San Diego just to work with us, and now you have become an inspiring leader and core contributor to KA Lite and FLE; I am very grateful that our paths crossed, and we now have a chance to work together, so I can continue to learn from you. Alexsandra McMahan, thank you for driving the FLE mission forward through helping us to spread the word, and supporting the team. Thank you to our remote contributors, especially Kriti Gupta and Steve Appelton, as well as many past (and, hopefully, also future) contributors who have had a big impact, particularly Vicky Tu, Matthew O'Rourke, Natalie Bui, Patricia Mou, Kian Lavi, and Arnaud Bénard. UCSD graduate students who have helped FLE and been an important part of the group include Jeremy Karnowski and Adam Mekrut. Thanks also to our fantastic board of directors, who have supported and guided our efforts: Bilal Musharraf, Marshall "Mike" Smith, Jeff Elman, Charles E. Rowe, Sarah Loos-Karnowski, and Catherine Hammack. And special thanks to the Hewlett Foundation and the Qatar Foundation International for financial support of our efforts, along with the UCSD Division of Social Sciences.

My parents, Karen and Eric Alexandre, have supported me, and my education, from before I was born until the present day, and I am forever grateful

to them for being such a big part of who I am and what I've been able to do. They are an inspiration to me, and have taught me so much, as they continue to do.

And last, but most definitely not least, I am so grateful to my dearest wife, Isabelle Tancioni. Meeting Isabelle was by far the best thing to have happened to me during grad school. Her advice, care, wisdom, and help have allowed me to be my best self, accomplish more than I thought possible, and enjoy life to the fullest. She has supported me through some of the lowest and highest points of my life, and with her by my side, I know I'll be able to face anything that comes my way.

Vita

- 2001-2006 Bachelor of Arts, Cognitive Science, Simon Fraser University
- 2007-2009 Master of Science, Cognitive Science, University of California,
San Diego
- 2007-2014 Doctor of Philosophy, Cognitive Science, University of California,
San Diego

ABSTRACT OF THE DISSERTATION

Computer Assisted Learning in a (Dis-)Connected Age:
Challenges and Approaches to Digital Education and Equal Access

by

Jamie D. Alexandre

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2014

Professor Jeffrey L. Elman, Chair

We are experiencing unprecedented global flow of knowledge and ideas, and seeing emergent opportunities for ad hoc educational experiences that are mediated by computer technologies and the Internet. Educators and learners are experimenting with new pedagogical approaches and modes of content delivery, both inside and outside of the classroom, in what has been described as an “online learning revolution”.

In this dissertation, I take a step back to examine the factors that are enabling this transformation, and demonstrate how not everyone has been benefiting equally -- in particular, recent developments have been failing to include many communities that are already among the most economically and socially disadvantaged. I present the work we have done to help bridge this gap, through designing, building, and deploying tools to allow learners in offline and bandwidth-constrained contexts to access the same types of novel educational opportunities as their more connected counterparts.

Two quite distinct lines of research and development are presented and discussed in this dissertation: an adaptive second language learning website, “ESL Genie”, and an offline deployment platform for Khan Academy, “KA Lite”. I will endeavor to demonstrate that these initiatives share a common emphasis on leveling the educational playing field by using technology to reduce economic barriers to effective learning experiences.

Chapter 1: Introduction and overview

What is the purpose of education?

The paradox of education is precisely this—that as one begins to become conscious one begins to examine the society in which he is being educated. The purpose of education, finally, is to create in a person the ability to look at the world for himself, to make his own decisions, to say to himself this is black or this is white, to decide for himself whether there is a God in heaven or not. To ask questions of the universe, and then learn to live with those questions, is the way he achieves his own identity. But no society is really anxious to have that kind of person around. What societies really, ideally, want is a citizenry which will simply obey the rules of society. If a society succeeds in this, that society is about to perish. (Baldwin, 1985)

Learning happens everywhere, both intentionally and as a byproduct of engaging in everyday life. Education, broadly interpreted, is when the process of fostering learning is formalized at larger scales, through the establishment of institutions and procedures, teachers and classrooms. Motivations for engaging in and supporting the practice of formal education, from the perspectives of both facilitators and learners, are naturally highly diverse, but a survey of historical and contemporary attitudes towards education reveals a number of distinct themes.

At a high level, the general goal of education could be thought of as passing the knowledge, skills, and traditions of one generation on to the next, whether in service of wisdom and enlightenment, economic competitiveness, or ideological propagation. The social goals and national goals behind formal education are often competing and even contradictory, as Baldwin identifies in the passage above. On the one hand, education has been championed and

utilized as a tool of indoctrination and obedience training, but on the other hand, education is also promoted as a means of developing the critical thinking skills that will enable an engaged and enlightened citizenry to question the practices and prejudices of their societies. From a socioeconomic standpoint, the goals and priorities behind formal education face a similar internal tension, with education often upheld as a powerful social equalizer, distributing opportunities and enabling social mobility, but in practice often also being exploited to perpetuate entrenched privilege and socioeconomic class.

In this dissertation, I will attempt to address these competing perspectives and motivations directly wherever possible, but the work described in the upcoming chapters is strongly driven by the goal of leveraging educational technologies to promote equal learning opportunities for both personal improvement and socioeconomic mobility, and by a recognition of access to education as a fundamental human right.

A brief history of educational inclusion

It may be tempting to try to paint a picture of the history of educational access and inclusivity as a natural progression along a continuum from elitist exclusivity to a modern era of enlightened universal education, but the reality is much subtler, geographically diverse, and non-linear.

For example, in ancient India from around 2000 BC, the early Vedic educational system -- which included areas such as grammar, composition, reasoning, medicine, astronomy, physics, and chemistry -- was egalitarian and

inclusive across social groups and genders, and the Vedic texts explicitly stated that anybody should be free to study the Vedic lessons. It was only later, Gupta (2007) argues, with increasing social stratification in the form of the caste system, that education became more discriminatory in India.

John Amos Comenius, a 17th century Moravian educator and theologian, has been referred to as the “Father of Modern Education” (Byers, 2000), and is considered one of the earliest campaigners for the cause of universal education. His “Didactica Magna” introduced such educational concepts as teaching 1) using practical illustrations and concrete objects, 2) following an increasing order of complexity derived from “natural progressions”, and 3) in an enjoyable and engaging fashion, in contrast to the rote pedagogy and dry presentation style of his era (The Great Didactic of John Amos Comenius, 1907). He advocated for education for *everyone*, “not the children of the rich or of the powerful only, but of all alike, boys and girls, both noble and ignoble, rich and poor, in all cities and towns, villages and hamlets” (The Great Didactic of John Amos Comenius, 1907). It is important to note, while his insistence that girls as well as boys be able to attend schools was indeed progressive for his time, in the face of hypothesized criticisms he reverts to a position that falls far short of universal education in any equal sense: “For we are not advising that women be educated in such a way that their tendency to curiosity shall be developed, but so that their sincerity and contentedness may be increased, and this chiefly in those things which it becomes a woman to know and to do ; that is to say, all that enables her to look after her household and to promote the welfare of her husband and her

family” (The Great Didactic of John Amos Comenius, 1907). However, along with other educational reformers to come, Comenius’ efforts to universalize education did help to push us one step closer to a world in which it could be a reality.

While support for the ideal of universal public education continued to grow through the centuries to follow, it was not until the United Nations was founded in the wake of the horrors of World War II that the principle of education being for everyone was enshrined into international law. The *Universal Declaration of Human Rights* (UDHR), Article 26 (UN General Assembly, 1948), states that:

Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit.

Since the signing of this international pact, most countries around the world have implemented some form of national policy aimed at ensuring a basic education is available to everyone at no cost. Levels of successful implementation of these policies vary widely, however, as we will outline below. While the UDHR does not specifically require that post-secondary education be made available for free, the *International Covenant on Economic, Social and Cultural Rights* (ICESCR), which was adopted in 1966 and went into effect in 1976, went one step further by affirming that “higher education shall be made equally accessible to all [...] by the progressive introduction of free education” (UN General Assembly, 1966).

Global barriers to equal educational access

Despite nominally having arrived at international consensus that education is a right and expectation for all, in the form of the UDHR and ICESCR resolutions described above, this is a battle that is far from being won. In 2000, a UN summit agreed upon 8 concrete but highly challenging objectives, known as the Millennium Development Goals, to be achieved by the year 2015. These included the target that “children everywhere, boys and girls alike, will be able to complete a full course of primary schooling” (United Nations, 2013). Through aggressive campaigns, partnerships, and funding, some progress towards this goal has already been achieved, with primary school enrollment rates rising from 83% to 90% between 2000 and 2011, and literacy rates rising. However, early dropout rates remain at the same 25% level as in 2000, and progress may now be faltering, as international aid for basic education slowed and actually began to fall in 2011. As of 2011, 357 million children are still not attending school, and over half of those are expected never to have the chance to do so (Rose et al., 2014).

The story of global educational inequality is a tragic story of compounding effects and vicious cycles. Very rural and poor areas, for instance, often lacking basic infrastructure, do not make attractive destinations for teachers, leading to a paucity of trained teachers and hence large student-to-teacher ratios in these areas where education could be of perhaps the greatest benefit (Rose et al., 2014). The parents of these poor and isolated children are also, first, less likely to be well-educated enough to teach their children themselves, and second, unable

to afford the private tutoring or extracurricular classes that could otherwise help their children succeed. With lower educational completion leading to lower expected lifetime earnings, these socioeconomic factors are then compounded further from generation to generation, leading to a cycle of poverty that we will not be able to break without proactive intervention.

Overview of the questions and challenges addressed in this dissertation

“When you can’t change the direction of the wind – adjust your sails.” – H. Jackson Brown

The topics in this dissertation dance a tango with the broad theme of educational technology and its role in boosting equitable educational access around the globe. We take the general approach of laying out the problems that we are attempting to solve, explaining the design of systems that may help to meet the needs that were identified, and then exploring the application of these systems within both experimental and real-world contexts. The work here operates at the microlevel, investigating subtleties of algorithms and technical approaches, as well as at the macrolevel, with attempts to build and deploy at scale systems that solve urgent global needs.

Chapter 2 explores work carried out in the field of intelligent tutoring systems and adaptive testing, applied in the context of an English as a Second Language learning website, ESL Genie. Much of the previous work in the adaptive language learning field has relied heavily on expert knowledge and hand-constructed curricular progressions; the approaches taken here are instead

very data-driven, while still built upon a solid foundation of manual work spent transcribing, annotating, and analysing ESL essays, to build a large corpus of multiple choice questions to use in the testing of experimental participants. This chapter begins with an overview of human tutoring, its effectiveness, and the role it plays in global learning equality, before turning to computer-facilitated tutoring and the field of second language learning. Item Response Theory (IRT) is explained and explored as a tool for diagnosing ESL learner proficiency, and then applied to the design and implementation of a website, ESL Genie, with which we then conducted user experience testing and a large battery of online tests, gathering as many measures as possible from over 400 participants (including detailed language background, hundreds of multiple choice quiz answers, a 300-word essay, and 3 blocks of self-paced reading). The data resulting from these experiments are then used as the basis for training and testing IRT algorithms, particularly the question-selection component aimed at selecting optimally diagnostic questions to optimize the efficiency of learner placement, followed by an examination of the relationships between the various measures collected during the experiments.

Chapter 3 delves more deeply into the role the Internet is playing in the shifts that are underway in global education, with the “online learning revolution”, and the barriers still in place that prevent most of the world, and in particular the most disadvantaged populations, from benefiting from them. It begins with a brief overview of the Open Educational Resource (OER) movement, followed by a survey of what has been happening in recent years with Massively Open Online

Courses (MOOCs), and the pedagogical innovations, such as self-paced learning and flipped classrooms, that are now being enabled. We then continue with a discussion of the digital divide, and its socio-infrastructure underpinnings, including a background on the emergence of the Internet, and the centralizing forces at play today. This will set the stage for the work described in the final chapter.

Chapter 4 describes the work we have done in attempting to address the educational and infrastructure issues described in previous chapters, though the design, implementation, and deployment of an educational software platform, KA Lite, that allows Khan Academy's diverse library of videos and interactive exercises -- along with user tracking, knowledge mapping, gamification, and coach reporting tools -- to be used without needing Internet connectivity. A background on the trends toward low-cost computing leads into the design considerations for making such a system feasible, and the goals we attempted to balance in implementing KA Lite. We then continue with a number of representative case studies of KA Lite in action in various contexts around the world. The chapter, and dissertation, then finish up with a survey of the linguistic diversity found in a number of OER repositories, and a discussion of future directions towards a platform for distributing additional OERs beyond Khan Academy, and enabling local teachers and learners to author their own content, bringing them in as equal participants in a global learning community.

Chapter 2: Tutoring, Computer Assisted Language Learning, and ESL Genie

Introduction

One-on-one tutoring plays an important role in the overall educational ecosystem, and the possibility of automated computer-mediated intelligent tutoring systems offers the potential to extend the reach and impact of tutoring to a larger audience. In this chapter, I will begin with an overview of some of the benefits that have been attributed to tutoring, and examine the issue of access with respect to who is currently able to take advantage of these benefits. I will then look briefly at some of the work that has been done to build intelligent tutoring systems using automated computer-mediated techniques, before diving more deeply into the application of these techniques to second language acquisition, in a field known as Computer Assisted Language Learning (CALL). This will set the stage for an overview of our design, implementation, and experimental testing of an English as a Second Language website, ESL Genie¹, and an exploration of the techniques we developed for analysing the results of these experiments, and their applicability to the construction of an optimally diagnostic quizzing system.

¹ <http://eslgenie.com>

Human and computer tutoring

The power of a tutor

Tutoring has been shown to be effective in multiple domains, including early reading (Elbaum, Vaughn, Tejero Hughes, & Watson Moody, 2000), mathematics (Hartley, 1977), and medical training (Groves, Régo, & O'Rourke, 2005). Engaging with a one-on-one tutor offers several clear advantages over large-classroom, lecture-style learning environments, including 1) greater opportunities for production and feedback, 2) the ability to engage in mediated self-paced learning through slowing down, speeding up, or digressing in order to best serve the learner's needs, 3) the ability to ask targeted questions, ala the Socratic method, to guide the learner to emergent insights.

In many areas of the world, access to a good tutor can make the difference between getting into a top school or a poorer school, with high stakes tests playing a large role in determining a large part of a student's success over the remainder of his or her life (Bray, 2006). For this reason, private tutoring and cram schools are big business, around the world, with the global market approaching \$100 billion annually ("Global tutoring industry experiencing explosive growth - ICEF Monitor - Market intelligence for international student recruitment," n.d.).

Equality of access to tutoring

As private tutoring typically falls outside the purview of the publicly funded educational system, it is not unsurprising that we can observe large social inequities in the ability to access and benefit from tutoring. In Turkey, for instance, 6.46% of the lowest quartile income families spent money on private tutoring, whereas 24.56% of the top quartile spent money on the same (Tansel & Bircan, 2006). The costs of tutoring are non-negligible, in some cases amounting to over 30% of a family's total monthly expenditures. These costs then feed directly into the spiral of social inequality, given the greater ability that groups with higher socioeconomic status have to leverage private tutoring to boost their children's educational levels, and aid them in passing exams with the high scores that will grant them access to more elite higher educational institutions and more profitable jobs (Bray, 2006).

The financial barriers that many families face in accessing the benefits of tutoring are further compounded in more rural contexts, where qualified tutors and remedial learning centers may simply not be available (Cheng, Liu, Ko, & Lin, 2007), but given the correlations between geography and socioeconomic status, it may in many cases specifically be the neediest students who are being left out. In areas that are being reached by the spread of the Internet, online tutoring services may help to fill in these gaps, particularly with the recent emergence of several nonprofit organizations that are operating websites to connect disadvantaged learners with qualified volunteer tutors over the Internet,

such as Learn To Be² and Tutor Chat Live Foundation³. While these efforts are helping many people, lack of access to the Internet, which is correlated with geographic isolation and low socioeconomic status, may pose a barrier to a large number of those who could benefit most. We will be exploring the educational barriers imposed by lack of Internet access further in Chapter 3.

Automated tutoring

Viewed within the context of unequal access to paid tutoring, and given the educational advantages this access can confer, developing successful, open systems for automated tutoring becomes more than just an intellectual exercise or neat technical trick -- it offers the potential for radically leveling the educational playing field, by lowering the barriers to entry by underprivileged students to the benefits of tutoring.

A successful adaptive tutoring system relies on several key components:

1. Some representation of the knowledge domain, whether specified as expert knowledge or inferred from a dataset
2. The ability to determine the state of the learners' knowledge; where are they in the landscape?
3. The ability to make helpful suggestions to the learner about next directions, based on a model of the connections and dependencies within the knowledge domain

² <http://www.learntobe.org/>

³ <http://www.tutorchatlive.org/>

4. The ability to adapt to a learner's choices and affective state, minimizing frustration and empowering the learner to choose what interests him or herself.

In this chapter, we will be focusing largely on #2, determining where in the knowledge space a learner is.

Computer Assisted Language Learning

Order of acquisition effects

It has been shown that successful human tutors tend to present materials according to a “curriculum script”, starting from easy tasks and moving through to challenging ones (VanLehn, 2011). In the domain of grammatical errors made by intermediate ESL students, which we are investigating here, it may be useful to first examine what is known about the sequential difficulty ranking and acquisition order of some of the morphemes with which the ESL learners are struggling.

The 1960s saw some of the first empirical investigations into the order of functional morpheme acquisition amongst children learning English as a first language. Roger Brown tracked the language use of three American children over a period of several years, and concluded that while the children mastered particular structures at different ages, the order in which they learned the structures demonstrated remarkable consistency across individuals (Brown, 1973). For instance, correct usage of “-s possessives” and irregular past tense (Stage III) always seemed to precede mastery of “-s third person present tense”

and regular (-ed) past tense (Stage IV). Brown suggested that adult learners of a second language might pass through comparable stages of acquisition, and subsequent research indicated that this may be partially true, although the order of acquisition differed somewhat from Brown's L1 stages; for instance, second language learners of English mastered "third person singular -s" before "-s possessives" (Dulay & Burt, 1973).

Early work on order of acquisition, inspired by Chomsky's notion of Universal Grammar, largely attributed these ordering effects to innate language knowledge or other "natural" constraints, citing consistent ordering across individuals from distinct language backgrounds (for example native Chinese and Spanish speakers) as they learned English (Dulay & Burt, 1974). Others sought to explain order effects in terms of inherent structural constraints and knowledge dependencies within the language itself:

"Sequences of development are as much, or even more, a consequence of epistemology, of the structure of knowledge in the relevant problem-space, as they are a consequence of the learners' biological processing capacity and neural development. Invariant developmental sequences of language acquisition are essentially interesting because they inform us about the informational content of language and how more complicated structures arise from simpler, more basic forms." (Ellis & Laporte, 1997)

There have also been more recent attempts to explain order effects in terms of more basic factors such as "perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency" (Goldschneider & DeKeyser, 2001), with limited success.

Order effects and pedagogy

What practical significance does the existence of consistent acquisition order have for the L2 learner and the L2 instructor? It has been frequently demonstrated that the order in which students acquire these morphemes is not reflective of the order in which they are being taught, and that students are relatively impervious to instruction targeted at an inappropriate stage of acquisition (e.g. (Lightbown, 1983)). Of course, consistent order of acquisition does not imply consistent rate of acquisition, and Krashen's influential "Input Hypothesis" (Krashen, 1982) suggests that students will be most receptive to input and instruction that lie just beyond the current ability level (referred to as "i+1").

Assessing a learner's knowledge state is already a difficult problem for educators, but needing to teach to a class of students with mixed backgrounds, strengths, and knowledge levels is a fundamental pedagogic challenge. Krashen's proposed solution in the context of second language education is to foster opportunities for interaction and communicative exchange, as this maximizes the variety of input that students receive, some of which will hopefully be at an appropriate level for pushing them forwards along their developmental trajectories. However, while there are clear benefits to be gained from communication-focused group interaction, we would also like to be able to provide students with more individualized input and feedback, to help them progress more efficiently. Unfortunately, this sort of personalized feedback can

necessitate private tutoring, which is out of reach for the majority of students, and many tutors may not have the knowledge or resources to effectively assess a learner's knowledge state.

Approaches to CALL

Efforts to make use of computational technologies in second language learning date back to the early days of digital computers. The appeal was obvious; a computer could present text to a learner and process the learner's responses, patiently repeating a task or an explanation as many times as the user required. The design of early systems was largely driven by a behaviorist approach to language instruction, involving drills and rote memorization, and these systems were implemented on mainframe computers that users accessed via a terminal (Warschauer, 1996). With the emergence of personal computers in the 1980s, development efforts in Computer-Assisted Language Learning (CALL) progressed in several new directions; for instance, advances in multimedia technologies allowed CALL systems to move beyond purely textual interfaces, and led to the production of rich audio-visual content, still a core component of most language learning software packages today.

Another line of CALL research sought to incorporate advances in the budding fields of Artificial Intelligence (AI) and Natural Language Processing (NLP) to provide more effective automated instruction and feedback to learners, an approach that became known as Intelligent Computer-Assisted Language Learning (ICALL) (for a review, see (Heift & Schulze, 2007)). The primary goals

of an ICALL system are to assess a learner's knowledge in order to provide appropriate feedback and tutorials, and to automatically detect and possibly correct erroneous learner productions. Typically this is carried out under fairly constrained conditions; the user is provided with a task, and the designer anticipates the types of errors the user is likely to make and implements mechanisms that allow these errors to be detected and reported to the user. For instance, a feature-based grammar (Hagen, 1994) may be coded to detect subject-verb number agreement, as in the following simplified example (ignoring articles):

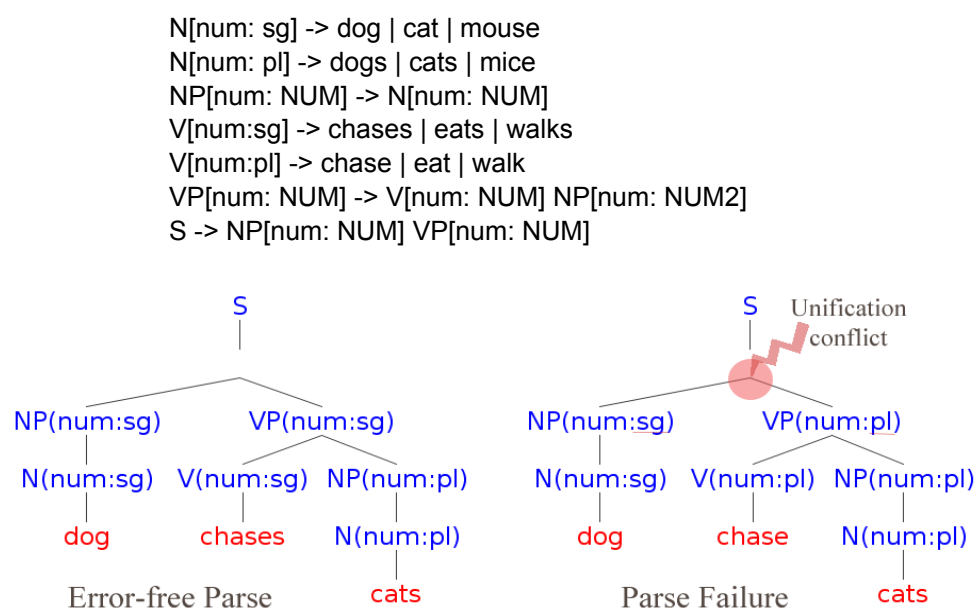


Figure 1: Examples of successful and failed parses in a unification grammar

In this grammar, lexical features (here, just grammatical number) are propagated upwards through the tree, and finally unified at the top level production, when the subject NP and predicate VP are joined together. In the

left-hand tree, the unification succeeds, as both the subject and predicate are singular, whereas in the right-hand tree, the unification fails, as the subject NP is singular but the predicate VP is plural. In a parser-based ICALL system, this failure to unify would result in the offending lexical items being flagged and an error message being displayed to the learner, possibly with links to further explanations or exercises related to the nature of the error.

One of the main problems with these sorts of ICALL approaches is the massive expert engineering effort that their development requires. While they are data-driven in the sense that the designers draw upon existing second language learners' text to identify error types, they require a large degree of manual intervention, and are thus not scalable. One reason for this is the lack of a suitable corpus of second language learner text; while a number of excellent L2 corpora have been gathered⁴, those that are publicly available do not come with corrected or annotated versions of the text, making them difficult to use as a basis for automated training of an error-recognition system.

Item Response Theory and methods for adaptive testing

Overview

Item Response Theory (IRT) is a popular approach to modeling the relationship between latent student proficiencies and the student's responses to testing materials. An Item Response Function (IRF) predicts the probability of a

⁴ e.g. The Uppsala Student English Corpus (USE) and Written Corpus of Learner English (WRICLE)

student answering a question correctly as a function of the student's underlying proficiency (or proficiencies, in the case of multidimensional IRT). A commonly used IRF is the 3-parameter sigmoid shown in Figure 2. The important parameters for our present discussion are a (the “discrimination” parameter) and b (the “difficulty” parameter). The larger b is, the “harder” the question is, in that a higher proficiency will be required to achieve the same accuracy level, and the higher a is, the steeper the slope of the sigmoid will be. The c parameter is also referred to as the “guessing” parameter, which is the probability of getting the question correct even when answering completely randomly, and serves as a lower asymptote.

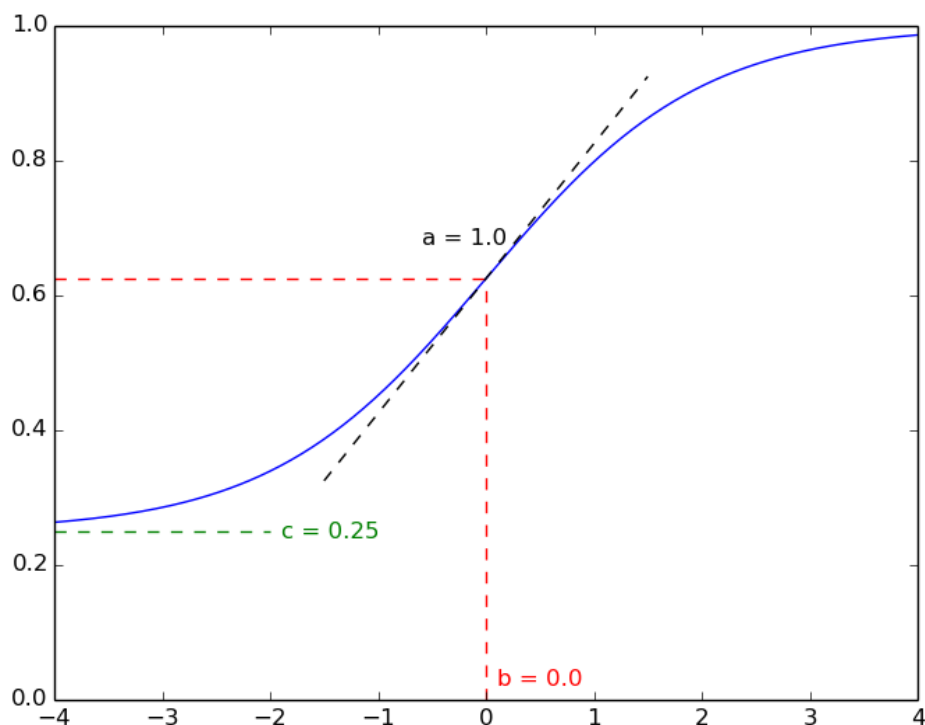


Figure 2: Example 3PL (3 parameter logistic) sigmoidal item response function

Structured knowledge domain

Though the current work will not progress to this stage (as more data would still need to be gathered), modeling the underlying structure of student knowledge could play a critical role in providing individualized and appropriate exercises and resources to students. Most models of the structure of student knowledge are hand-built by a researcher, in the manner of an expert system, rather than taking full advantage of a corpus of user data in combination with machine learning techniques. Ideally, we could assume only the very high-level organization of the knowledge domain, and have the actual structure of the student model be induced from a large corpus of student response data. One way to approach this is by using a Bayesian belief network (see Figure 3 below) with a layer of unobserved “knowledge component” nodes that represent aspects of a learner’s knowledge (in our example here, grammar in a second language); the meaning of these nodes need not be specified in advance, but one might imagine them corresponding to beliefs such as “subjects need to agree in number with verbs”, or “adjectives should precede the nouns they modify”. There are directional connections in the graph between these hidden nodes, representing dependencies between knowledge components; these connections too would be learned on the basis of user responses. A number of learner parameters, such as overall proficiency and language background (e.g. age of exposure, the learner’s native language, and time spent studying other languages), would set a prior on the value of the hidden nodes. Proficiency levels

within each of the knowledge components then specify a distribution over learner responses (e.g. answers to multiple choice questions).

One of the advantages of coding the learner's knowledge in this kind of Bayesian belief network is the flexibility it allows in the types of inference that could be made. For instance, if the parameters could be trained to predict mastery of the various Knowledge Components on the basis of a learner's L1, we could then use Bayes' rule to make the reverse inference (assuming we have assigned prior distributions, by means of hyperparameters, on Language Background), and estimate a posterior distribution over the learner's L1 on the basis of their performance.

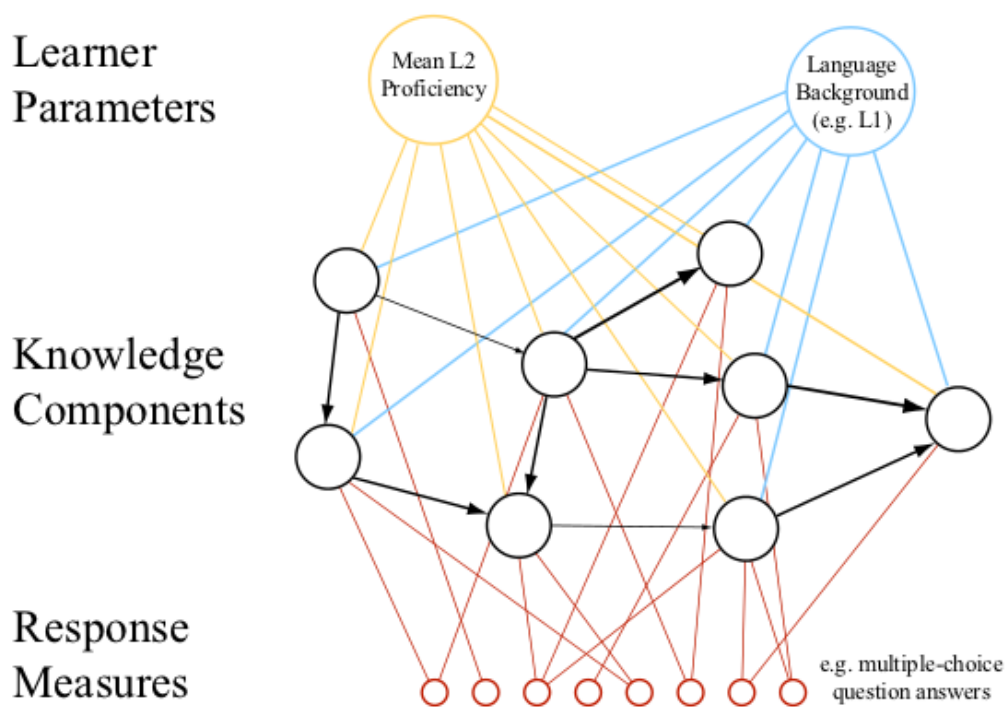


Figure 3: Hypothetical graph structure used to predict response measures

Inference in a Bayesian IRT system

For simplicity in the following explanation, we will assume a single hidden knowledge component (latent proficiency variable), which puts a distribution over the responses to each of the observable quiz questions (Figure 4). For now, to simplify our demonstration, let us assume that the prior distribution on the latent variable is a Gaussian, with a mean of 0, and the distribution on each observable node is a 2-parameter IRT sigmoid (with parameters learned from the data).

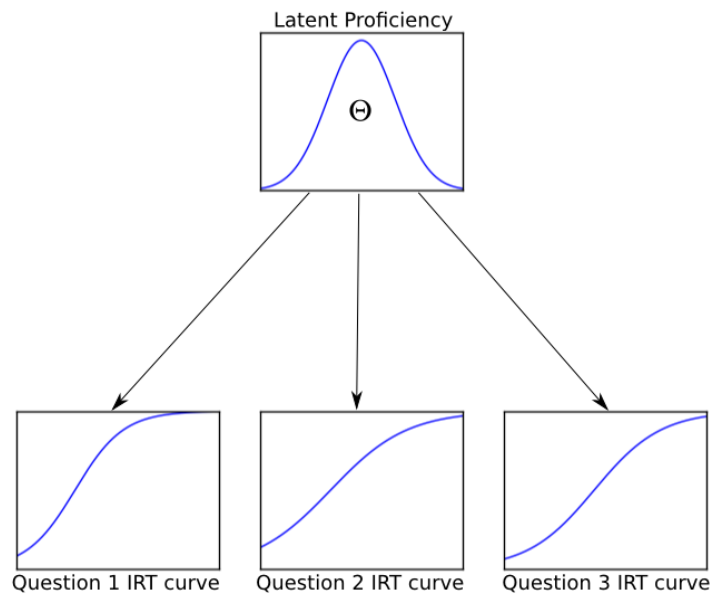


Figure 4: Distributions and links between latent proficiency and response measures

C_{ij} : indicator variable, equal to 1 if person i is correct on question j ,
otherwise 0

θ_i : person i 's proficiency (expressed in standard deviations from the mean)

a_j : IRT parameter a (slope of sigmoid, aka “discrimination”) for question j

b_j : IRT parameter b (horizontal shift of sigmoid, aka “difficulty”) for question j

There are several types of inference we would like to make within this model. Firstly, assuming we have an estimate (represented as a Gaussian distribution) of the learner's proficiency level θ_i , we may want to predict his or her accuracy on a particular response measure. This is a straightforward calculation, involving passing the proficiency distribution forwards through the question's sigmoid:

$$P(C_{ij}|\theta_i, a_j, b_j) = \theta_i \cdot \text{sigmoid}(a_j, b_j)$$

The only difference at this point with standard IRT is that instead of assuming θ_i to be a single value, we allow it to be a full distribution, and hence instead of evaluating the sigmoid at a single point, we take the dot product of θ_i and the sigmoid to obtain a weighted estimate of the learner's probability of getting the question correct. The next type of inference we would like to make is of a learner's proficiency on the basis of his or her responses to the multiple choice questions, where we can take advantage of Bayes' rule:

$$P(\theta_i | C_{i1..n}, a_{1..n}, b_{1..n}) = \frac{P(C_{i1..n} = 1 | \theta_i, a_{1..n}, b_{1..n}) \cdot P(\theta_i | a_{1..n}, b_{1..n})}{P(C_{i1..n} | a_{1..n}, b_{1..n})} \propto \prod_j P(C_{ij} = 1 | \theta_i, a_j, b_j) \cdot P(\theta_i)$$

Between the second and third steps of the derivation above, we take advantage of the fact that the correctness values $C_{i1..n}$ are conditionally independent of one another conditioned upon the proficiency score θ_i , as are the IRT parameters $a_{1..n}$ and $b_{1..n}$. The elegant result of this simplification is that the posterior probability over a learner's proficiency can be calculated by simply multiplying the prior probability $P(\theta_i) = \text{Gaussian}(\mu = 0, \sigma = 1)$ by the IRT sigmoid curves of the questions answered correctly, and one minus the IRT sigmoid curves of the questions that were answered incorrectly.

When we examine the effect of a student's question answers on our posterior estimate of his or her proficiency, a few patterns emerge (see Figure 5). The blue dotted line shows the prior probability estimate of the learner's proficiency, which is a Gaussian distribution with mean of 0 and standard deviation of 1. The red lines show a variety of item response functions for questions of increasing levels of difficulty, and the solid blue lines show the posterior distribution resulting from correct and incorrect answers to those questions.

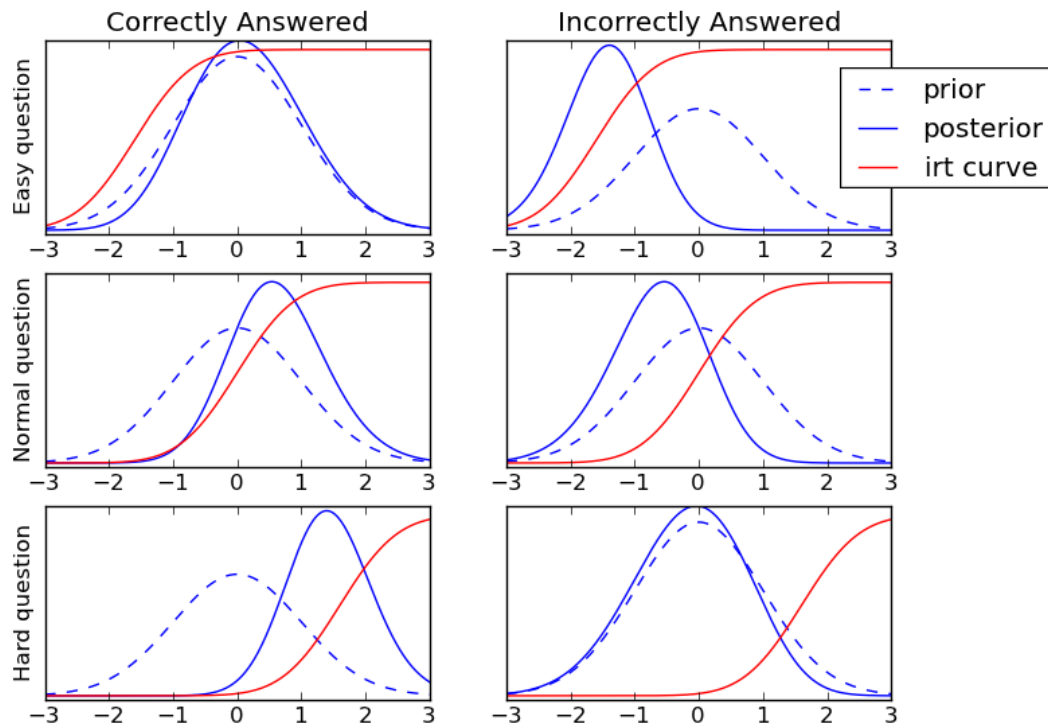


Figure 5: Effect on the posterior proficiency estimate of answering questions with varying difficulty either correctly or incorrectly

Figure 6, below, provides a different view onto the same pattern, showing the results of 5000 simulated answers for answers with a broad range of difficulties and slopes (“discriminations”). The height of the points is the mean of the posterior distribution (i.e. our resulting estimate of proficiency), and the colors of the points indicate the standard deviation of this resulting estimate (red is higher variance, and blue is lower variance). Again, we see that answering hard questions incorrectly or easy questions correctly has little effect on our estimate of proficiency, whereas answering easy questions incorrectly reduces proficiency, and answering hard questions correctly increases our proficiency estimate.

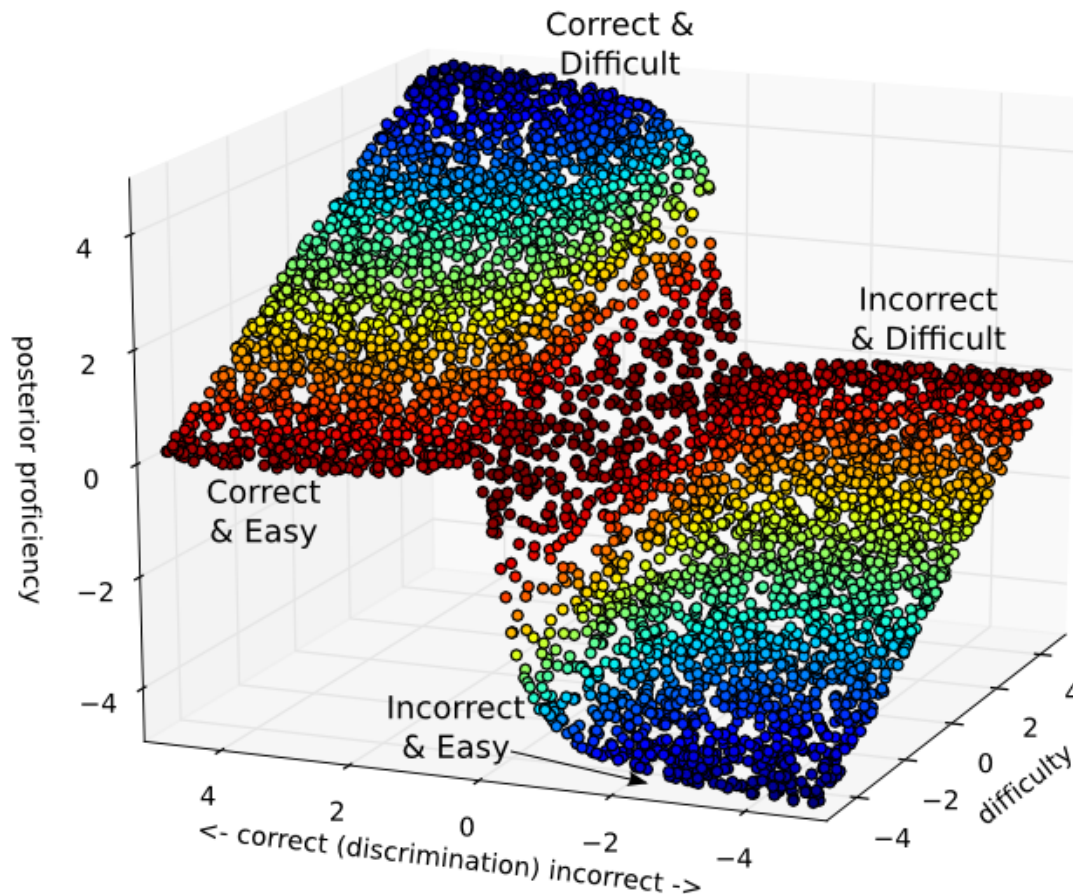


Figure 6: Effect on mean estimated proficiency of question difficulty and discrimination (cooler colors mean a lower-variance, i.e. more confident, posterior)

ESL Genie: An investigation into adaptive error diagnosis in ESL learners

Introduction

When we began what was to become the ESL Genie project, our first task was to create the multiple choice question bank that would be used for gathering data in the experiments. We wanted the grammatical errors tested by these

questions to be a broad and accurate reflection of the type of errors made by our target population (college-level English as a Second Language learners from diverse linguistic backgrounds), and so we needed a good corpus of advanced ESL text to use as a source. After scouring dataset listings, we found several candidate datasets, but found them each to be lacking in some respect; in particular, the level of the language was generally either too low (written by very beginner learners), or too homogeneous (written by learners in a particular foreign country). Fortunately, we were able to partner with the Basic Writing Program at UC San Diego to obtain a large corpus of writing samples from the Analytical Writing Placement Exam (AWPE), a handwritten exam administered to incoming University of California undergraduate students who had not completed English 12 or achieved sufficiently high scores on the writing portion of their SATs. Of the 1420 exams we obtained, 175 were marked as being from English as a Second Language learners (half of whom were US citizens, and around a quarter of whom had lived in the US their entire lives). We did not have access to their first language data, but were able to guess the approximate proportions of various language backgrounds on the basis of surnames (see Figure 7 below).

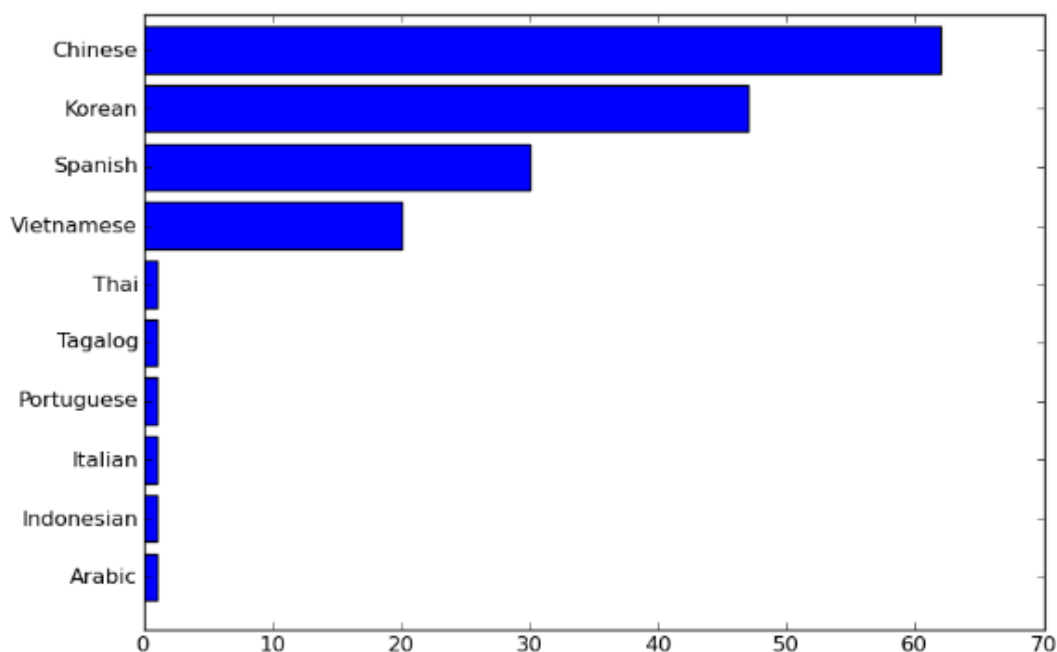


Figure 7: Estimated distribution of first languages in the AWPE essay corpus

In total, we transcribed and annotated 81 ESL essays (52,375 words), and 51 non-ESL essays (28,519 words), developing a coding scheme (see Appendix A) as we went along, to mark grammatical and other errors inline in the transcribed text, both for later analysis and to use in building questions for the test bank. Due to the frequently convoluted handwriting, we added a “magnifying lense” feature to aid in text inspection.

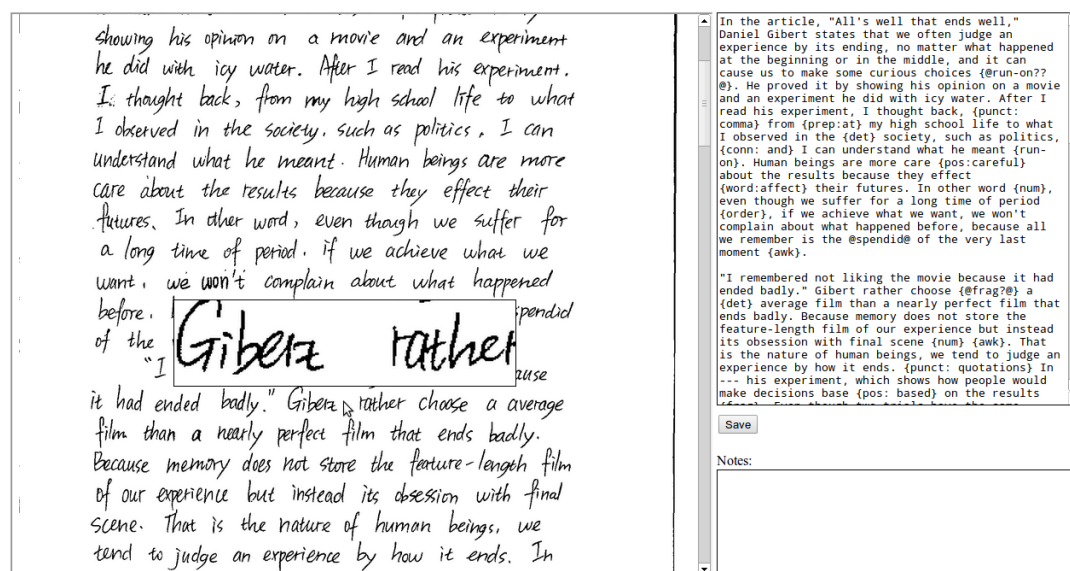


Figure 8: AWPE transcription interface with magnifying glass.

After the transcription process was well underway, we began brainstorming multiple choice questions that could try to capture some of the patterns of errors we observed in the student-written essay text. Many of the questions were derived directly from sentences in these essays, while others were created from scratch but using the error categories identified during the process of annotation.

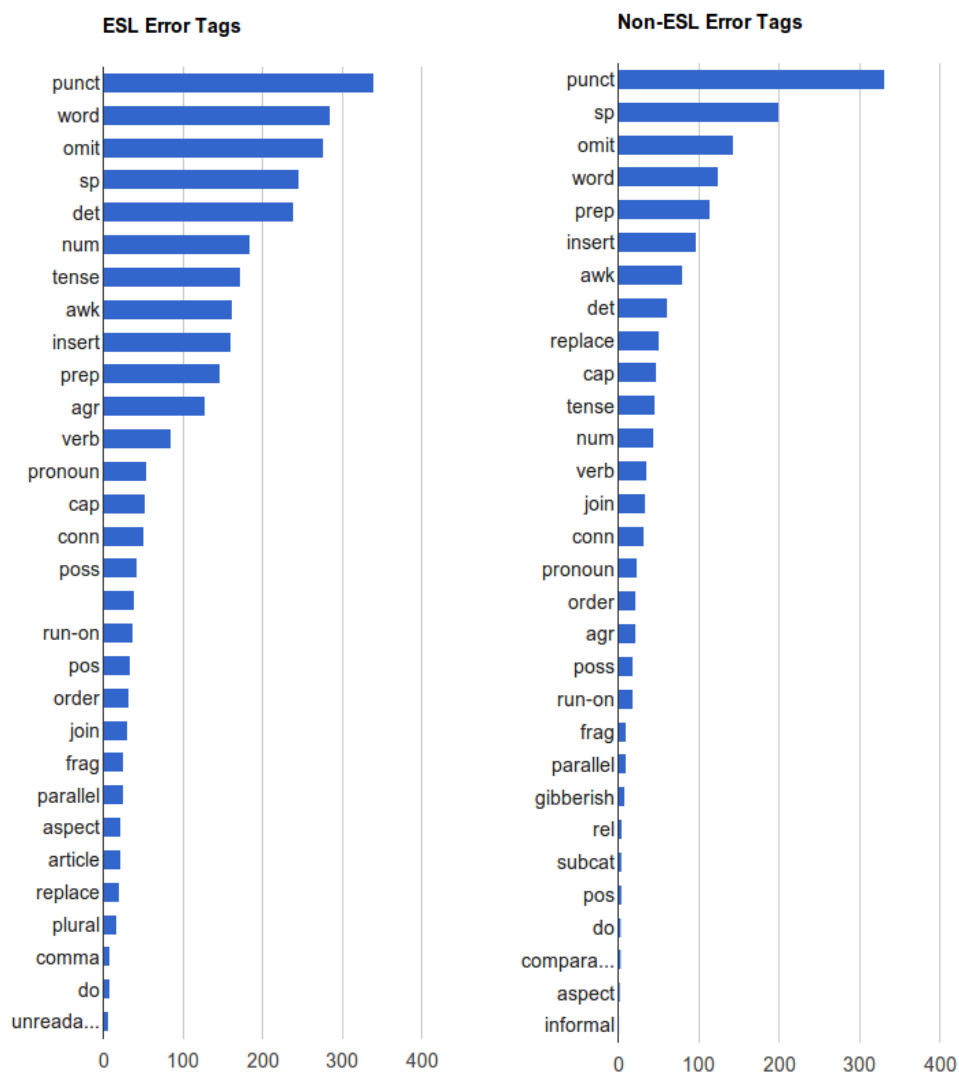


Figure 9: Distribution of error tags in the transcribed essay corpus from ESL and non-ESL writers

Quiz interface design

Once we had a question bank underway, a lot of care went into designing the quizzing interface, in attempting to simultaneously satisfy the following sometimes competing goals:

1. *Keep the user flowing through the questions.* As we wanted to gather as much data as we could, and keep learners engaged, we didn't want users to get bogged down in a particular question, or have to click too many times to continue.
2. *Provide clear and immediate feedback.* Following a response, the user should immediately know whether the provided answer was correct, and if not, what the correct answer was.
3. *Avoid discouraging the user.* We want to be able to indicate to the user gently, and in an encouraging manner, when he or she makes a mistake, to avoid frustrations, and pencils being jammed into monitors.
4. *Allow the user to easily review the question just completed.* Without interrupting the flow or losing state, the user should be able to review the previously completed question.
5. *Avoid distracting the user, by keeping the focus on the current question.*
6. *Reduce cognitive load.* We need the user's working memory free to dedicate to the actual content of the question, so that we're testing the right thing.

To aid with goal #1, we took inspiration from *freerice.com*, which has users click on a matching vocabulary word, and then immediately proceeds to the following question, showing terse feedback on the previous response in a line above the new question. We took this one step further in order to meet goal #4, using a "card" metaphor where the new question card slide would slide in on top of the old question card, while leaving the old card partially visible (though faded,

to reduce salience and thus distraction, to serve goal #5). To allow the user to review the previous question easily (goal #4), the partially obscured and faded card would temporarily slide back out and brighten if hovered, allowing for easy review without any clicks or departures from the current context.

To help with #2, we focused on immediately changing the answer colors, following a click, to indicate the correct (green) and incorrect (red) answers.

Initially, we only changed the color of the answer that was clicked and of the answer that was correct (in some cases, these were of course the same).

However, during user tests and interviews, we saw some confusion about how to interpret these colors, especially in cases where the question involved identifying the phrase with an error in it. To one user, it was unclear whether the red was indicating that the phrase was the incorrect one, or that it had incorrectly been selected. To help with goals #2 and #6, we changed the behavior to have *all* of the incorrect answers turn red, the correct answer still turn green, and the selected answer (whether correct or incorrect) become bolded and underlined (see Figure 9, below). As subjects soon learned that all the questions only have one correct answer, the contrast between the green answer and the consistently red answers seemed to help with disambiguation.

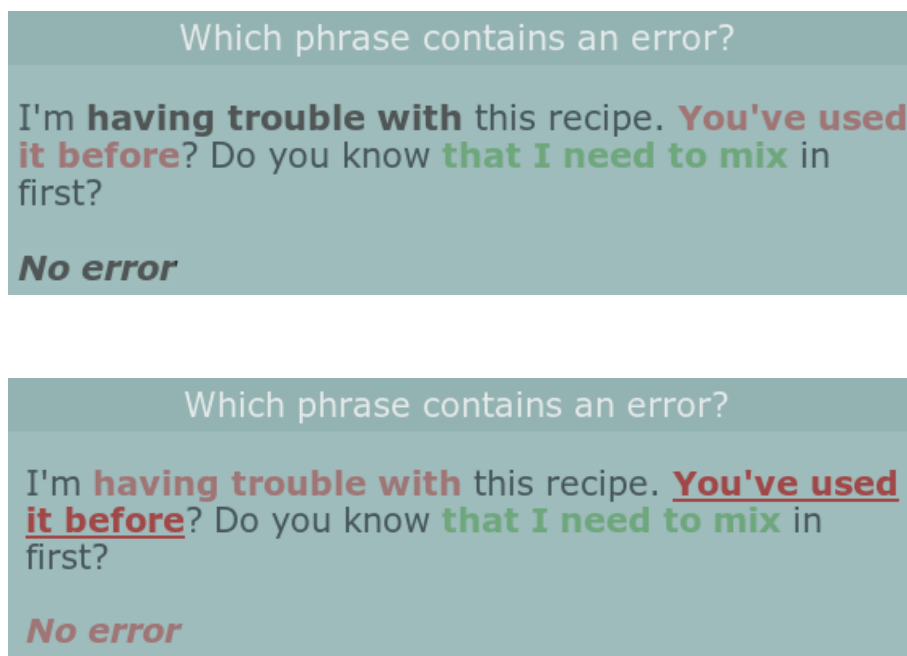


Figure 10: Versions of color-coding correctness on questions, before user testing (top) and after user testing (bottom)

As additional feedback to the user after clicking a response, particularly as the previous card moves up and behind the new card fairly rapidly, we implemented an overlaid check mark (for correct) or X (for incorrect), which faded in and out briefly during the card transition. We went through several iterations of the “incorrect” overlay, in attempting to reduce its “negativity” (to aid in goal #3), but in the end we realized that it was the outcome of getting the question wrong, rather than the mode of feedback, that was associated with the negative responses. Once the tests were made adaptive, we expected these frustrations to be reduced, as a long string of incorrect answers would be less likely given the auto-calibration of question selection towards the user’s level.

For goal #6, reducing cognitive load, an approach we implemented for “fill in the blank” questions was to have the response text for the currently hovered answer appear inline in the gap in the question body, allowing it to be read in context without the visual and cognitive overhead of context switching between the answer list and the sentence (see Figure 11, below). During UX testing, one user exclaimed that they loved this feature, and found it really useful.



Figure 11: Inline “fill in the blank” question display on hover

Methods

Subjects were recruited from the UC San Diego undergraduate population, and participated for 3 hours of class credit. The experiment listing requested participants for whom English was not their first language, though many had lived in the United States their entire lives. Subjects were provided with an account on ESLGenie.com, and logged on to begin the experiment.

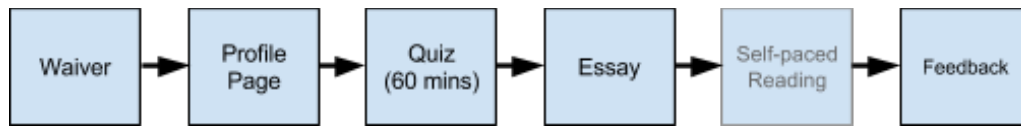


Figure 12: Subject flow through the stages of the experiment.

After completing a waiver and providing informed consent, they were directed to their profile page, which they were required to complete before proceeding. The purpose of the profile page was to gather some background information that could be relevant to the subject's knowledge of English. First, it gathered information specifically about the subject's history with, and self-reported ability for, the English language. Subjects could also specify the English variant with which they were most familiar, through an auto-completing search box that pulled language data from Ethnologue (Lewis, Simons, & Fennig, 2013). Self-assessments for reading, writing, listening, and speaking were requested independently, using qualitative descriptions that were aligned with a numeric scale.

English Background

English variant: Canadian ☒ Native speaker

When did you start learning English? 2

How many years have you used English? 4.0

In what environments have you used English the most?

☐ Home, with family/partner

☒ School, language courses

☐ Primary language in school

☒ Lived in country of language

☐ Primary language in workplace

☒ Other (describe below)

Reading: Fluent **Writing: None**

Listening: Intermediate **Speaking: Advanced**

None Can't understand anything

Beginner Can understand a few key words

Intermediate Can understand general meaning of speech

Advanced Can understand almost everything, except a few key words

Fluent No problem understanding complex, rapid speech in most topics

Native Identical to a native listener

that can't be captured above.

writing and speaking? (optional)

Figure 13: English background information on the profile page

Subjects were then asked to add any other languages they know, beginning with their native language. To ensure standardized inputs, and to elicit specificity around the regional dialect the subject spoke, a drop-down list of matching languages was displayed as users typed in the names of their languages (again, drawing from the Ethnologue database of 7,105 known living languages).



Figure 14: Language autocomplete tool to gather detailed information about languages spoken

After subjects finished listing all known languages, we needed some way to allow them to intuitively and efficiently specify an estimate of the proportion of time they spent using each of the languages over the past year. After experimenting with various approaches, and in the process of conducting user tests and interviews, we converged on a method involving slider bars, with a dynamically updating pie chart indicating the proportions of each language, shown below. These efforts paid off in encouraging users to provide an accurate estimate of their language usage, with the following observation found in the UX testing journal: “Being very detailed with the pie chart -- interesting. Yay for fun GUI’s!”

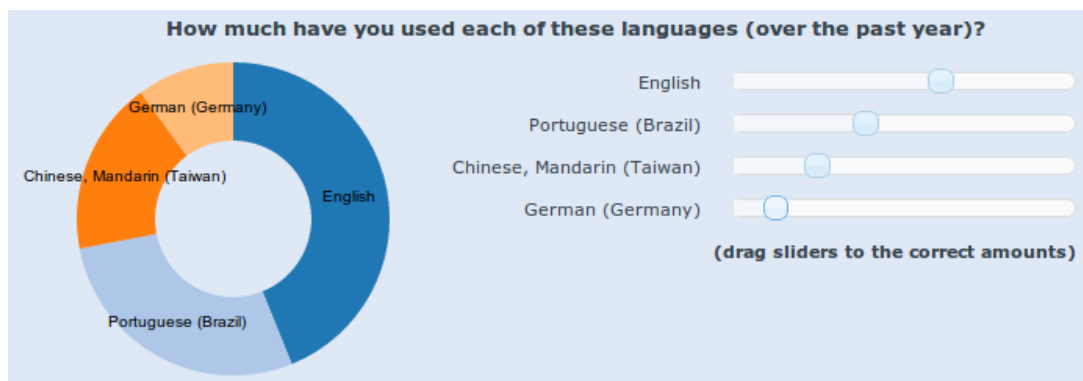


Figure 15: Language proportion interface

After completing and saving the profile page, subjects were redirected to the quizzing page, where they had 60 minutes to complete as many questions as they could (but with no incentive to rush, and warnings if they seemed to be going too fast). Questions were presented using the quizzing interface described above, allowing subjects to move along quickly and answer a large number of questions. After all 347 questions were completed, or 60 minutes (minus breaks) had elapsed, the subject was directed on to the essay-writing portion of the experiment.

**Please write a short story describing the picture
(min. 200 words)**



Figure 16: Image prompt for the essay-writing component.

To encourage free text production for use in error analysis and syntactic complexity measurement, subjects were presented with the image shown above, and asked to write a short story about what was going on in the scene. For the second half of the experimental subjects, the text editor used was Etherpad⁵, which tracks every insertion and deletion, along with their precise timing, such that the writing history can later be replayed in real time, or analysed through automated techniques.

After completing their essays, the first batch of subjects was directed to a feedback page where they debriefed and were assigned their class credits. For the second batch of subjects, an additional self-paced reading task was inserted here first, to provide data that would allow us to investigate possible correlations between reading times and question performance. Three separate stories were

⁵ <http://etherpad.org/>

presented one at a time, with all the words converted into underscores, and replaced with the actual word one at a time as the user pressed the spacebar. The time spent on each word was recorded.

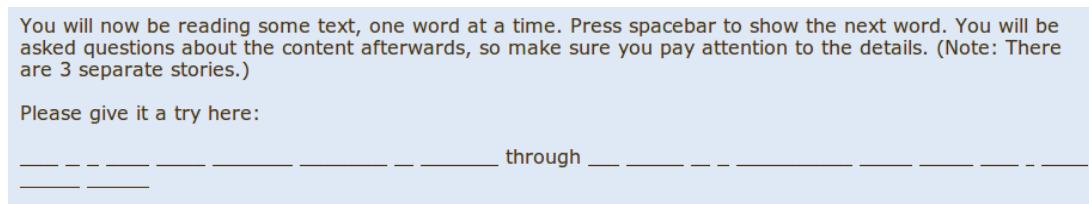


Figure 17: Introduction to self-paced reading task

After completing each of the stories, subjects were asked a series of short comprehension questions, to keep them focused on processing the sentences and determine when they had not been paying attention.

Exclusion criteria

We expected some portion of the subjects not to take the experiment seriously or pay attention (particularly following one research assistant's shocking admissions regarding to the techniques he typically used to get through experiments with the least effort possible), and so we included two experimental measures for the purpose of applying exclusion criteria, allowing us to remove the obvious bad data prior to analysis. First, we included 8 “dummy” questions in the pool of questions, which were then inserted into the question stream at intervals throughout a session. These questions were constructed such that any

careful reading by the subject population should leave no ambiguity as to the correct answer (see examples below, in Figure 18).

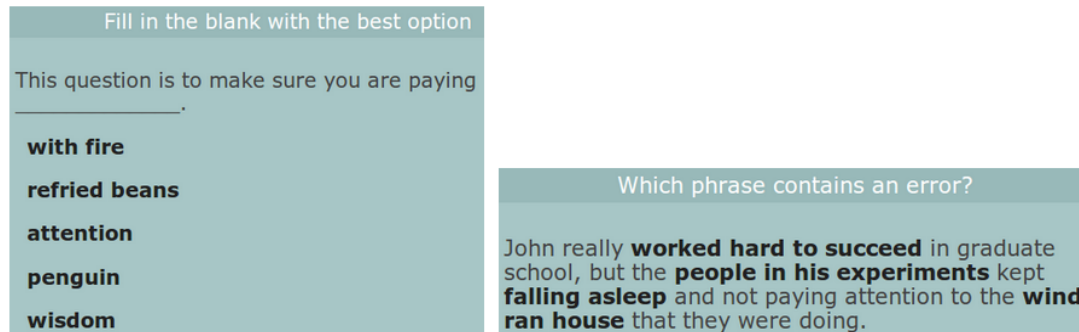


Figure 18: Examples of dummy questions, designed to detect subjects who were not attending to the task.

Secondly, we were recording response times for each question, and knowing the approximate lower bound on the amount of time needed to read the questions, we could detect rapidly clicking subjects and exclude them from analysis as well.

Demographics

Given the ESL requirement included in the experimental listing, we expected a broad set of linguistic backgrounds, but were surprised at just how diverse the native languages of our participants were (see Figure 19, below). As in the language estimates for the AWPE essays described above, Spanish, Chinese (here broken into multiple dialects), Korean, and Vietnamese were the

most commonly seen first languages, but with a long tail of less common languages.

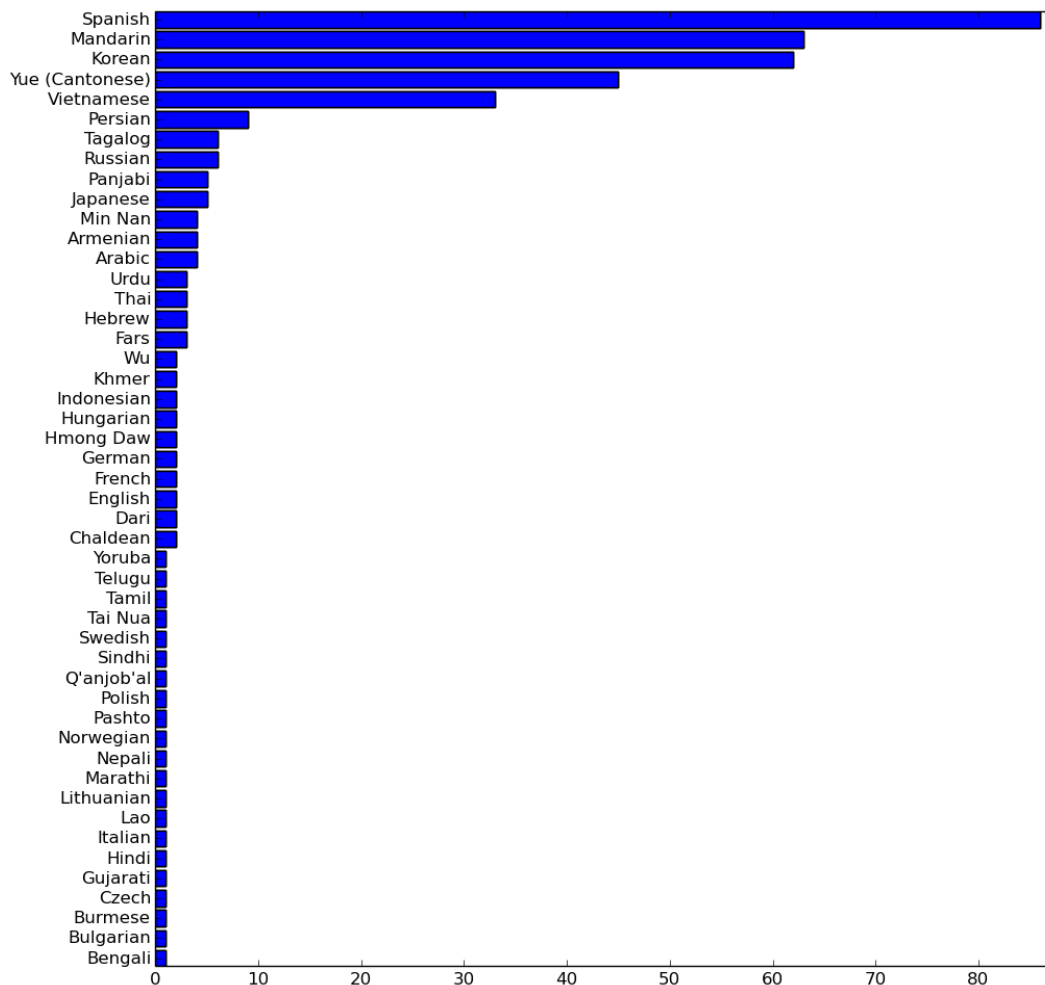


Figure 19: Language backgrounds of included participants

Analysis

First, we applied the criteria described above to determine subject exclusion. While the majority of subjects seemed to put in a good effort to answer the questions carefully and at a reasonable pace (many also commenting at the

end that they enjoyed the process and felt they learned something), the data for a number of subjects indicated that they were attempting to get through as quickly as possible without paying attention (and one subject even appears to have made use of mouse automation techniques, with a substantial portion of question responses less than 300ms). By examining the proportion of questions answered correctly as a function of the time taken to answer them, we can observe that there is a quick dropoff in accuracy below around 2.5 seconds (see Figure 20), indicating that these responses are frequently not good estimates of subject ability, and may largely consist of random clicking. Subjects were excluded if more than 21.5% of their responses (2 standard deviations above the mean) were faster than 2.5 seconds.

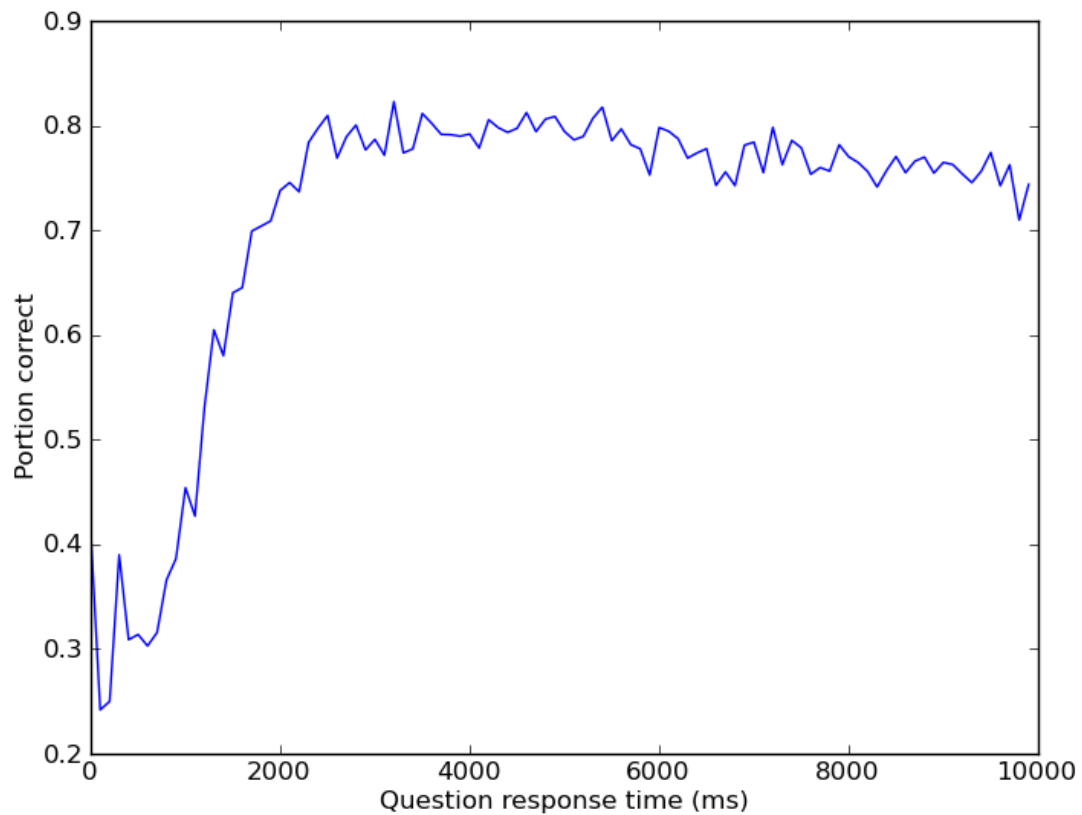


Figure 20: Correctness as a function of response time

Subjects were also excluded on the basis of incorrectly answering a large portion of the dummy questions described above. On average, subjects answered over 90% of these questions correctly, but subjects with less than 62.8% dummy question accuracy (2 standard deviations below the mean) were excluded from analysis. In the end, out of 435 subjects, 16 subjects were excluded for being too fast in their responses, 22 were excluded for incorrectly answering too many dummy questions, and 8 were excluded by both criteria, for a total of 46 exclusions (10.6%).

Questions with greater than 95% overall accuracy were also excluded from analysis, as they provide very little diagnostic information about participant ability levels, and can cause problems for convergence of the algorithms. On this basis, 31 questions, along with the 8 dummy questions, were excluded, leaving a total of 308 questions.

IRT parameters were learned for these 308 questions using the *mirt* package for R⁶, which made use of Expectation-Maximization to infer the optimal parameters. The model type chosen was the 4-parameter logistic (4PL), which includes upper asymptote and lower asymptote (chance) parameters in addition to the 2PL slope and difficulty parameters.

Results from Item Response Theory testing

The first thing that can be observed when we look at the trained Item Response Function parameters, by means of their sigmoidal curves graphed in Figure 21 below, is their great diversity; this large variance in the shifts and slopes of the curves for the questions in our test bank bodes well for our ability to diagnose a wide range of abilities, as we will be able to select questions as needed to be optimally diagnostic based on our evolving estimate of a subject's proficiency.

⁶ <https://github.com/philchalmers/mirt>

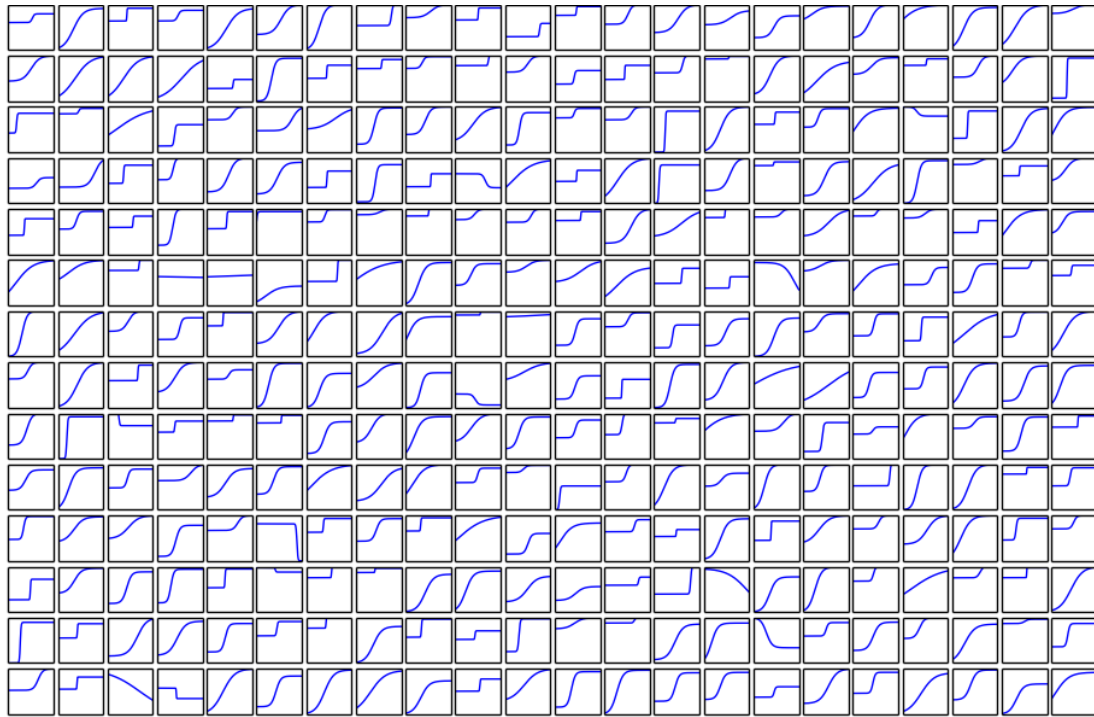


Figure 21: Item response function curves for all of the 308 included questions, each graph having an x domain of $(-4, 4)$ and y range of $(0, 1)$.

It may be worth taking a moment to look at a few of these questions and their curves in a bit more detail, as this will help to build up the reader's intuitions for interpreting the upcoming discussion and analysis making use of Item Response Theory. Recall that the Item Response Function maps from a subject's underlying latent proficiency, θ , to a probability of that subject answering the question correctly. A curve going up and to the right (as in the "Difficult" curve in Figure 22 below) means that as a subject's proficiency increases, the probability of answering that question correctly also increases. The fact that this curve is shifted somewhat to the right is why we have labeled it "Difficult", as compared

with the curve for the “Easy” question, which has a high probability of being answered correctly by all but the lowest proficiency subject.

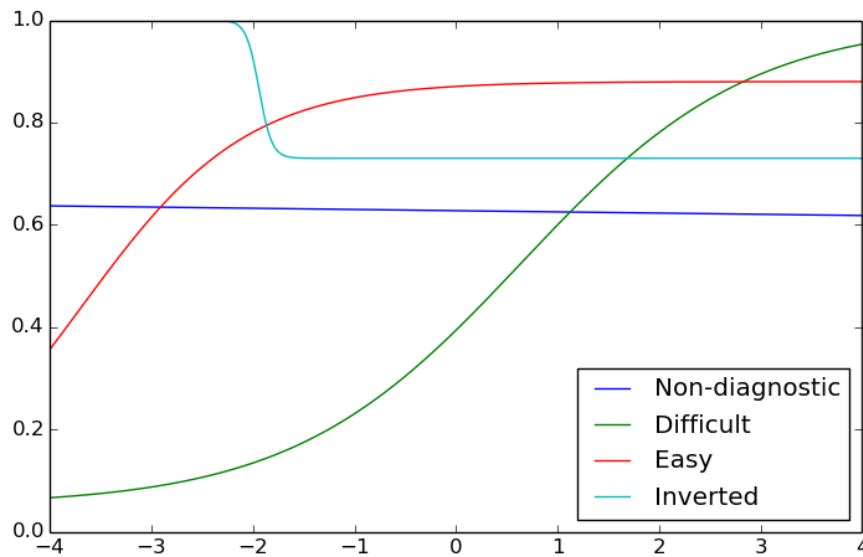


Figure 22: Exemplar curves from questions in our test bank.

The “Inverted” case deserves special attention, as its negative slope could be a warning sign that the wrong answer has been marked as correct in the question bank. What this negative slope effectively tells us is that better subjects actually do worse on this question than the lower proficiency subjects. An example of a question like this is shown below in Figure 23, where a possible explanation is that higher proficiency subjects are “overriding” the double “-ing” based on prior knowledge, versus the lower proficiency learners who are using more deliberate strategies.

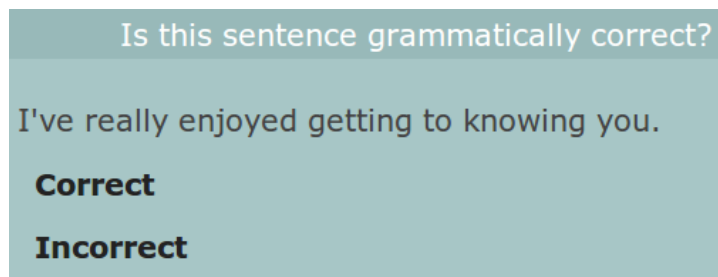


Figure 23: Question with negative IRF slope, more often answered correctly by lower proficiency subjects.

The “Non-diagnostic” curve, on the other hand, is unredeemably unuseful. Being flat at 0.6 across the board means that subjects of any proficiency, when faced with this question, have a 60% chance of answering correctly. Thus, whether a subject answers the question correctly or incorrectly, we have learned nothing about the subject’s actual underlying competence. An example of this type of question is shown below.

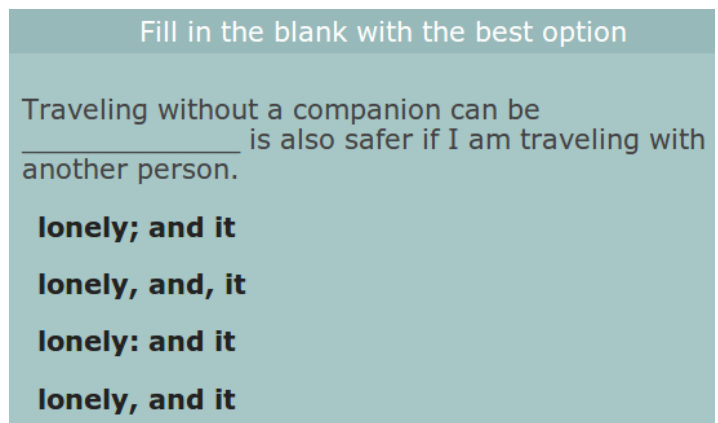


Figure 24: A non-diagnostic question, characterized by being answered incorrectly equally often by the top performers and the bottom of the barrel.

This brings us now to the question of diagnosticity. In the experiments we conducted here, we wanted an unbiased estimate of the IRT parameters, and wanted subjects to answer as many questions as possible, so we did not concern ourselves with optimizing the efficiency of the question ordering to converge as quickly as possible on an accurate estimate of a subject's ability.

As examples of potential (but not globally optimal) strategies, consider the following. For those with a high level of underlying ability, choosing the most difficult questions first leads to quicker convergence on an estimate of their true ability, as it allows them to “prove themselves” on questions that most people would have gotten wrong. Conversely, for those with low underlying ability, starting with the easiest questions allows us to hone in more quickly on an accurate estimate of their ability, as they will likely get even these easy questions wrong, which would be very unexpected for those of higher ability.

One commonly used strategy for optimal ability diagnosis is *entropy minimization* (or “expected uncertainty reduction”), which aims to choose questions that will maximally increase our posterior confidence about the subject's underlying ability. We can formalize this notion in the following procedure, applied to each available question in turn:

- Estimate the subject's probability P of getting the question correct, by passing the distribution θ , representing our current estimate of the subject's ability, through the IRT sigmoid for that question.
- Estimate the posterior distribution θ_c that will result if the subject answers the question correctly.

- Estimate the posterior distribution θ_i that will result if the subject answers the question incorrectly.
- Calculate the expected entropy (uncertainty) of the subject's ability, by weighting the entropy of the possible outcomes by their probabilities:

$$P \cdot \text{entropy}(\theta_c) + (1 - P) \cdot \text{entropy}(\theta_i)$$

- Selecting the question with minimum expected entropy as the next question to present to the subject.

A distinct, but related, question selection algorithm is *information maximization*, which places a priority on questions that are maximally discriminating (i.e. have the steepest slope on the IRT sigmoid) at our current estimate of the subject's ability (θ). Formally, this can be computed by multiplying the distribution of our current posterior estimate of θ by the derivative of the sigmoid curve for the question, as this gives us a θ -weighted estimate of the level of discrimination the question is likely to have.

Entropy minimization and information maximization both seem to be efficient strategies for question selection from our question bank. The average error (measured as the KL divergence of the final posterior ability estimate from the estimate at every timepoint) is much lower than with random question selection (see Figure 25, below). The “lower bound” error shown in the figure is a theoretical limit, using an ideal sequence for each subject that is calculated to minimize the *actual* error at every timepoint, by always selecting questions that will bring the posterior estimate as close as possible to the final posterior ability distribution (of course, since this involves advance knowledge of how the subject

will respond to future questions, this selection algorithm cannot be used in practice).

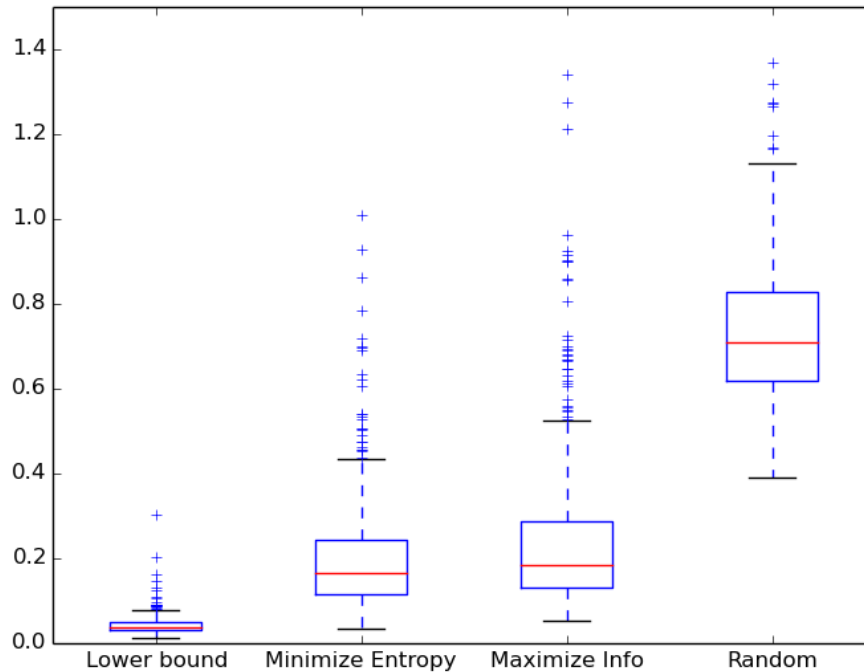


Figure 25: Average error (across subjects) of leading question selection algorithms

While the means of the errors for entropy minimization and information maximization are fairly close (0.210 vs 0.257, respectively), this difference is highly significant ($p < 0.001$ in a two-tailed t-test), meaning that for this question bank and population, entropy minimization comes out ahead in terms of optimal ability diagnosis.

It is also interesting to compare the similarity between the question sequences themselves, as generated by each of the selection algorithms. These sequences, of course, vary quite widely across subjects, since the choice of

question at every timepoint is highly influenced by the history of responses, which determines our current posterior estimate of the subject's ability. To test the similarity between sequences both within and across selection algorithms, we ran a series of simulations using the following procedure:

1. For each selection algorithm:
 - (a) For each subject:
 - i. While unasked questions remain:
 - A. Choose the next question to “ask” according to the selection algorithm (based on our current estimate of the subject's ability, and the IRT parameters of the available questions), then check how the subject actually answered that question to update our posterior estimate of the subject's ability.
2. Calculate the similarity between each pair of question sequences using an approximation of the “edit distance”⁷, normalized to fall between 0 (no similarity) and 1 (identical).

Using this procedure, we can compare sequences across subjects within a selection algorithm (see the left-hand side of Figure 26 below), or across selection algorithms (either for different subjects, or for the same subject).

⁷ Using the “ratio” method on the SequenceMatcher class from Python's difflib library.

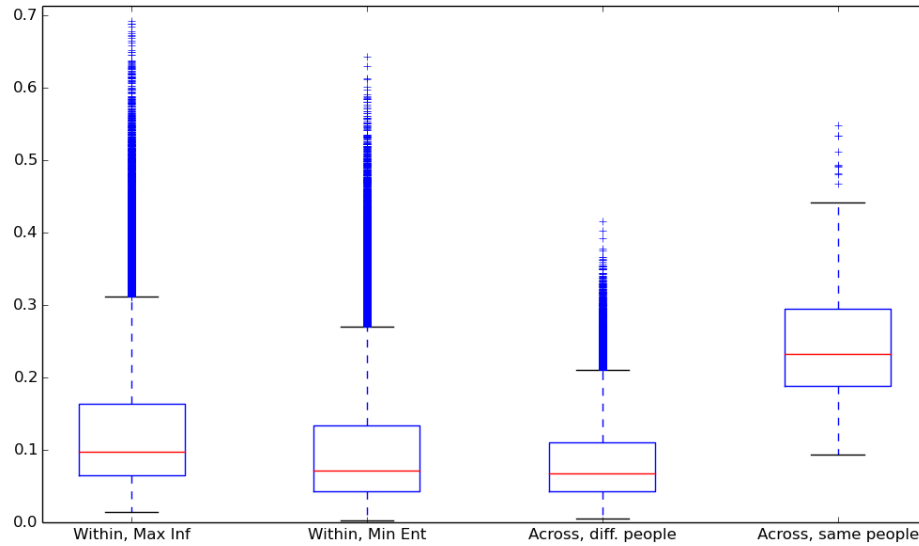


Figure 26: Similarity between sequences generated by different question selection algorithms

A few interesting things can be observed from these comparisons. Firstly, the sequence similarities are overall very low, on average (recall that the similarity metric ranges from 0 to 1). Secondly, the right-most column, which represents comparisons between the sequences generated by the maximum information and minimum entropy algorithms, for the *same person*, is substantially higher than the others. What this means is that the subject characteristics are a much stronger determinant of the selected sequence than the choice of algorithm.

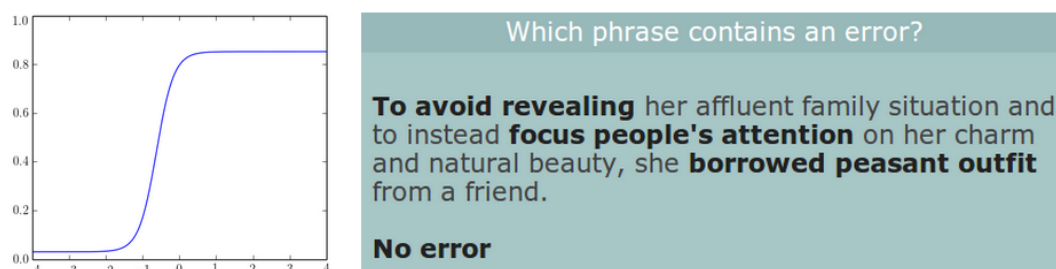


Figure 27: IRT curve and question text for Q85, picked as the first question by both entropy minimization and information maximization

Exploring information sources for estimating prior ability

For all the simulations up to this point, we have been using a Gaussian prior probability distribution with a mean of 0 and standard deviation of 1, reflecting our lack of prior knowledge about the subject's ability level. However, we actually have several pieces of information that could allow us to construct an informed prior estimate of subjects' ability levels, particularly in the form of their self-reported English proficiency. First, however, we need to examine the relationship between subjects' self-ratings of their own proficiency, and our posterior estimate of their ability on the basis of their question responses. Figure 28, below, shows this correlation to be surprisingly strong, giving some hope that these self-ratings could be used as the basis for a more informed prior estimate of ability.

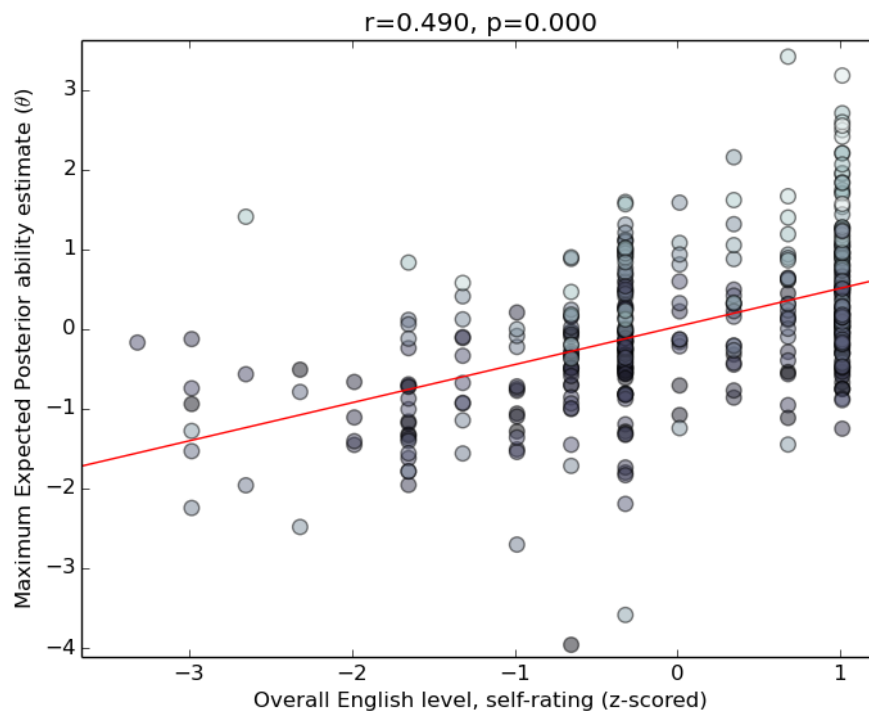


Figure 28: Relationship between self-rated proficiency and posterior estimate of actual ability

To explore the usefulness of these self-ratings in the process of adaptive diagnosis, we tested using the self-reported overall English ability, z-scored to bring it into the same scale as θ , for the mean of the Gaussian prior distribution (instead of the default mean of 0). Note that it would be complicated to estimate the effect this has on overall error in an unbiased way, as by changing the prior we also change (sometimes dramatically) our eventual posterior ability estimate. For this reason, the following investigation was conducted in an exploratory fashion, and the following cases are meant to serve as illustrative examples of the potential advantages and pitfalls of this technique. The posterior estimate

traces, along with error and uncertainty of the estimates, are shown below in Figures 29 and 30, with the standard mean of 0 used in the left-hand plots, and the subject's z-scored self-reported ability used as the mean in the plots on the right.

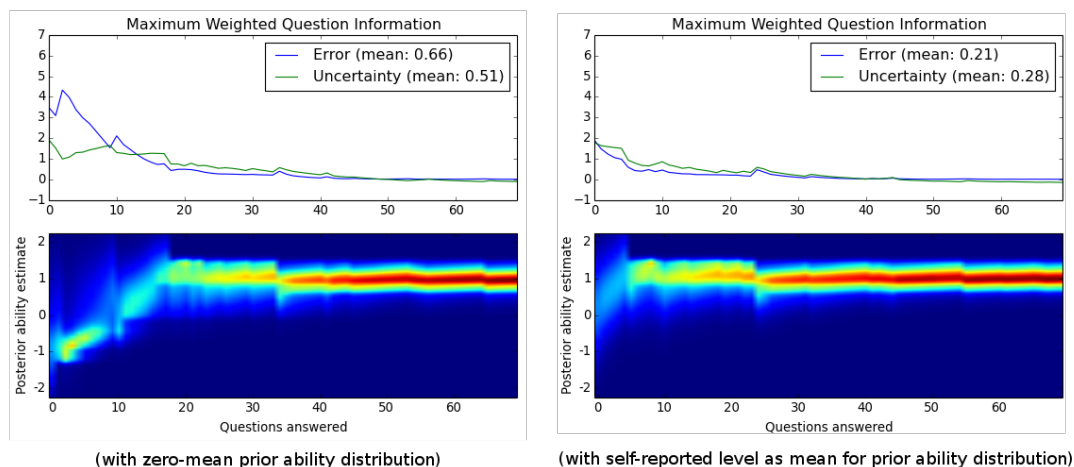


Figure 29: Effect of using self-reported proficiency as the prior mean for subject #93, who accurately provided a high self-assessment (self-rating of 0.826, posterior θ of 0.815)

Subject #93, shown above, estimated his or her own English language ability as moderately high, which coincided closely with the model's posterior ability estimate. For this subject, starting the prior probability distribution with a mean of 0.826 (as seen on the righthand plot), led to much faster convergence on a confident estimate of the true ability than when the prior mean started at 0. Notice the slower upwards slope of the posterior on the lefthand graph, only reaching a confident estimate after 35 questions have been asked, as compared with the graph on the right, where the confidence already starts to increase after about 25 questions. This provides some hope that we could indeed improve our

diagnostic efficiency by pre-seeding the prior with the subject's self-reported levels.

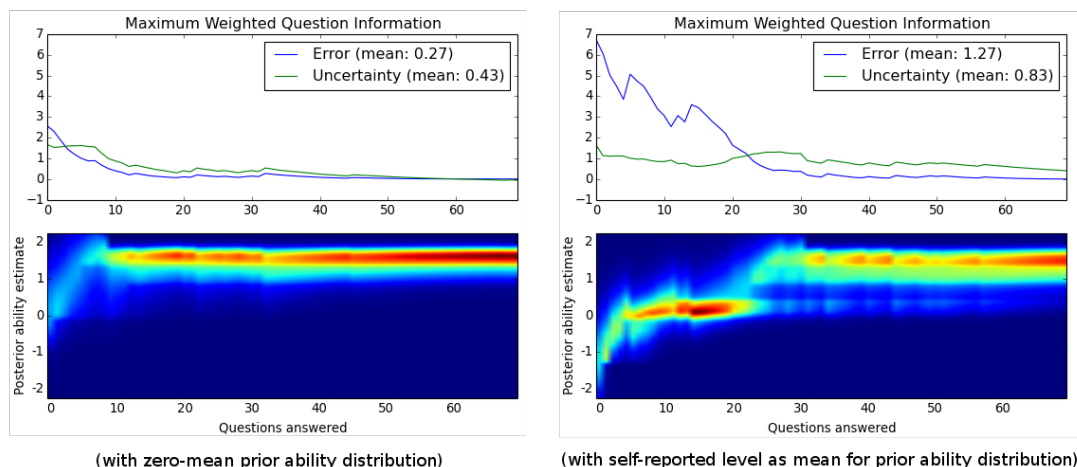


Figure 30: Effect of using self-reported proficiency as the prior mean for subject #71, who was overly modest in self-reporting ability (self-rating of -1.791, posterior θ of 1.409)

Unfortunately, the results are not always as compelling as we saw with subject #93, particularly when a subject's self-reported proficiency is not corroborated by his or her performance in the testing. A particularly dysfunctional case of this can be observed with subject #71, shown above, whose self-reported proficiency (-1.791 standard deviations below the mean) was dramatically lower than their actual performance (with a posterior θ estimate of 1.409). Convergence in the default case, on the left, actually happens quite rapidly. When we start the prior mean at -1.791, on the other hand, not only does it take much longer to converge on a correct estimate of θ , but it actually first converges on quite a confident but *incorrect* estimate, from questions 5 through 25, before self-correcting after it has gathered sufficient evidence.

In this section, we have seen that there is potential value in incorporating prior knowledge about a subject, such as self-reported ability levels, into our prior distribution of proficiency, but that great care must be taken when doing so that these information sources are reliable and do not bias the prior unproductively. One of the striking things that can be observed in all of the above figures, however, is just how quickly the adaptive quizzing process does tend to converge, and that it *can* actually recover from a badly biased prior distribution quite efficiently -- in the examples above, after asking 30 or so questions (around 10% of the total bank of 308 questions).

Results from the self-paced reading task

The goal of including the self-paced reading task alongside the multiple choice quizzing and essay writing components of the experiment was to see whether we could find relationships between our other measures of language ability and this popular technique for measuring implicit sensitivity to word frequency and grammatical violations. In the self-paced reading paradigm, a subject reads through a series of sentences where all the words except one are masked, and the subject presses a key to proceed from word to word at a self-selected pace. The time spent on each word reflects a combination of a large number of factors, including word length, word frequency, position in the sentence, and also grammatical violations, which tend to lead to slower response times in the region immediately following the violation. Differences between self-paced reading results for native and non-native speakers have been

reported in the literature (Jiang, 2004; VanPatten, Keating, & Leaser, 2012), with the general observation being that for many types of violations, non-native speakers do not show sensitivity as reflected in their self-paced reading times.

As described above, we presented 204 subjects with 3 different bodies of text of varying genres, in a masked self-paced reading paradigm. The text was not originally designed to contain explicit violations, but instead several types of grammatical structures that less-advanced ESL learners are known to have difficulty processing, such as relative clauses (Izumi, 2003). However, as these structures did not at first investigation appear to show any strong results, it was our good fortune to have also accidentally included a tense error on a verb in the first story, and it is the results from this region of text that we will focus on for our analysis here (see the sentence in italics and the bolded error word in the paragraph below).

There was once a father penguin and his son Pablo sleeping in their home in the Antarctic. After the father penguin woke up his son was nowhere to be seen. The father penguin discovered that Pablo had left a note on the table. In the note, Pablo said that before returning to the Antarctic from his stay at SeaWorld, he had fallen in love with the daughter of a polar bear who enjoyed making snow angels. Pablo had decided to sneak onto a plane that he knew would return to San Diego, where he would be able to get back to SeaWorld. The father penguin was not mad because he wanted his son to be happy with the one who he learns Pablo loves. He also knew that San Diego is pretty nice and he would be able to go visit and relax with Pablo on the beach.

Pablo arrived in San Diego the next day. *He snuck out of the plane at the airport, and **realize** that he had to go find a taxi that could take him to SeaWorld.* Pablo eventually got a taxi driver's attention, and they drove off to SeaWorld. During the ride, Pablo told the driver about how he met and fell in love with a polar bear named Patricia as he looked outside at the scenery. It made the driver so

happy that he stopped enjoying the scenery, focused on driving, and gave Pablo the ride for free. Once Pablo arrived at SeaWorld and found Patricia, he explained how he travelled all the way from his home in the Antarctic so they could spend the rest of their lives together. She was not excited by this news and told Pablo how she had a new love interest. Pablo was distraught by the discovery of another individual who Patricia believes loves her. On the bright side, Pablo's trip made him realize that there is so much available in this world. He then vowed to forget about Patricia and travel the world to discover what life has to offer.

Following the observations in the literature, our expectation would be that this violation would be detected more strongly by native speakers than by second language learners of English, and thus that native speakers would show a larger slowdown in the region following the violating verb. Technically, our subject pool in this experiment was intended to be purely ESL, but there was clearly a broad spectrum of ability, and some subjects were definitely “more native English-speaking” than others. To examine the relationship between subjects' proficiency scores (θ) estimated from the multiple choice quizzing responses as described above, and the impact on their reading times of encountering the grammatical violation, we first split off the very high performing ($\theta > 1.5$) and very low performing ($\theta < -1.5$) subjects, and took the within-group median of their response times (the time from when the word was shown to when they pressed the spacebar to proceed) for each of the words in the sentence containing the error. The pattern, shown below in Figure 31, strongly matches our expectation that subjects with higher English proficiency would react more by slowing down following the syntactic violation than the lower proficiency subjects. Not only is this result significant ($p < 0.05$, and often $p \sim 0.01$) across most of the

words in the window following the error, but the effect size is also staggeringly large, in the order of 100s of milliseconds (almost a doubling of response time).

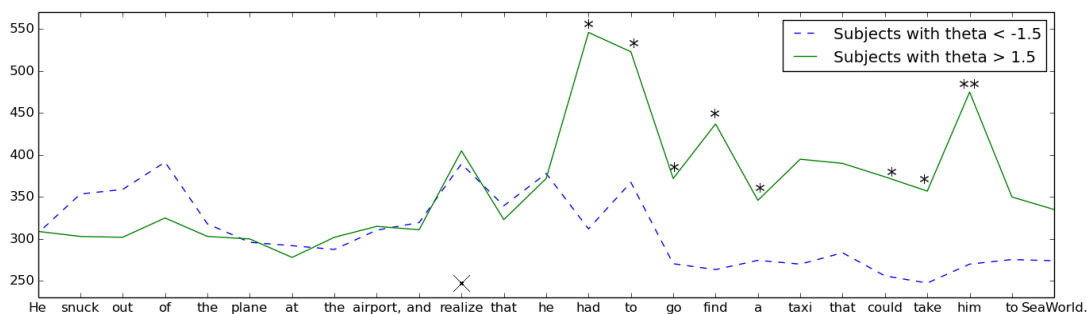


Figure 31: Median self-paced timings (in ms) across the sentence with error at “realize”, binned by low and high proficiency

Curious as to whether this was more of a bimodal effect (only showing up in the extremes) versus a graded effect (where the response slowdown varies continuously across the proficiency spectrum), we performed the same analysis, but this time with subjects binned into 4 different groups on the basis of their proficiency (Figure 32, below). A clear and beautiful pattern emerges, with the slowdown effect growing larger and larger as proficiency increases across the groups.

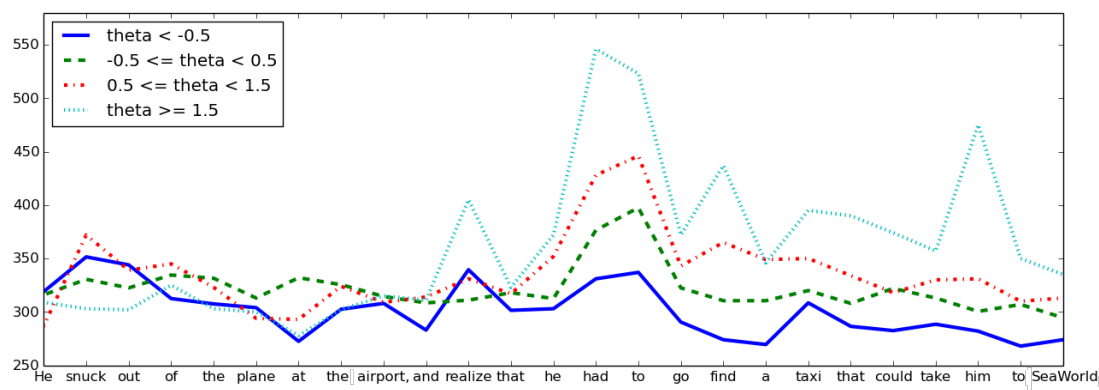


Figure 32: Median self-paced timings (in ms) across the sentence with error at “realize”, binned into four different groups by increasing proficiency

The strength of this relationship indicates there is great potential for adaptive language tutoring systems to benefit from including self-paced reading as a diagnostic activity, to provide insight into subjects’ implicit⁸ grammatical competencies in a potentially less affectively charged paradigm than repeated questioning where there are “correct answers”.

To further explore the relationship between our posterior ability estimate θ and the subjects’ reading times, we calculated a single mean time per subject across the words from “realize” to “go” (inclusive), and then performed a scatterplot and Pearson’s r computation with θ , which turned out to be highly significant ($p < 0.01$). Interestingly, though we saw earlier that θ is highly

⁸ It is possible that there was a substantial explicit component to the responses to grammatical violations as well, particularly since all the subjects had just completed an hour of answering questions that involved grammatical scrutiny and error identification, and may therefore have been primed for this; one subject commented afterwards that “It was a strange story, and coupled with the first part of the experiment I was mainly focused on the grammar of the story”. Of the 204 subjects who completed the self-paced reading and were asked “Was there anything confusing about this text?”, 12 mentioned having noticed grammatical errors, some specifically mentioning issues with tense. These subjects, who were almost all in the upper ability range, did have slightly higher reading times around the peak at “had” than their peers of equal mean proficiency, supporting the possibility that explicit attention may have played a role for some subjects.

correlated with the subjects' self-reported ability levels, these self-reports have no correlation at all with the reading times, suggesting that there is a strong independent source of information in our estimate of θ that could not have been derived a priori (see Figure 33, below).

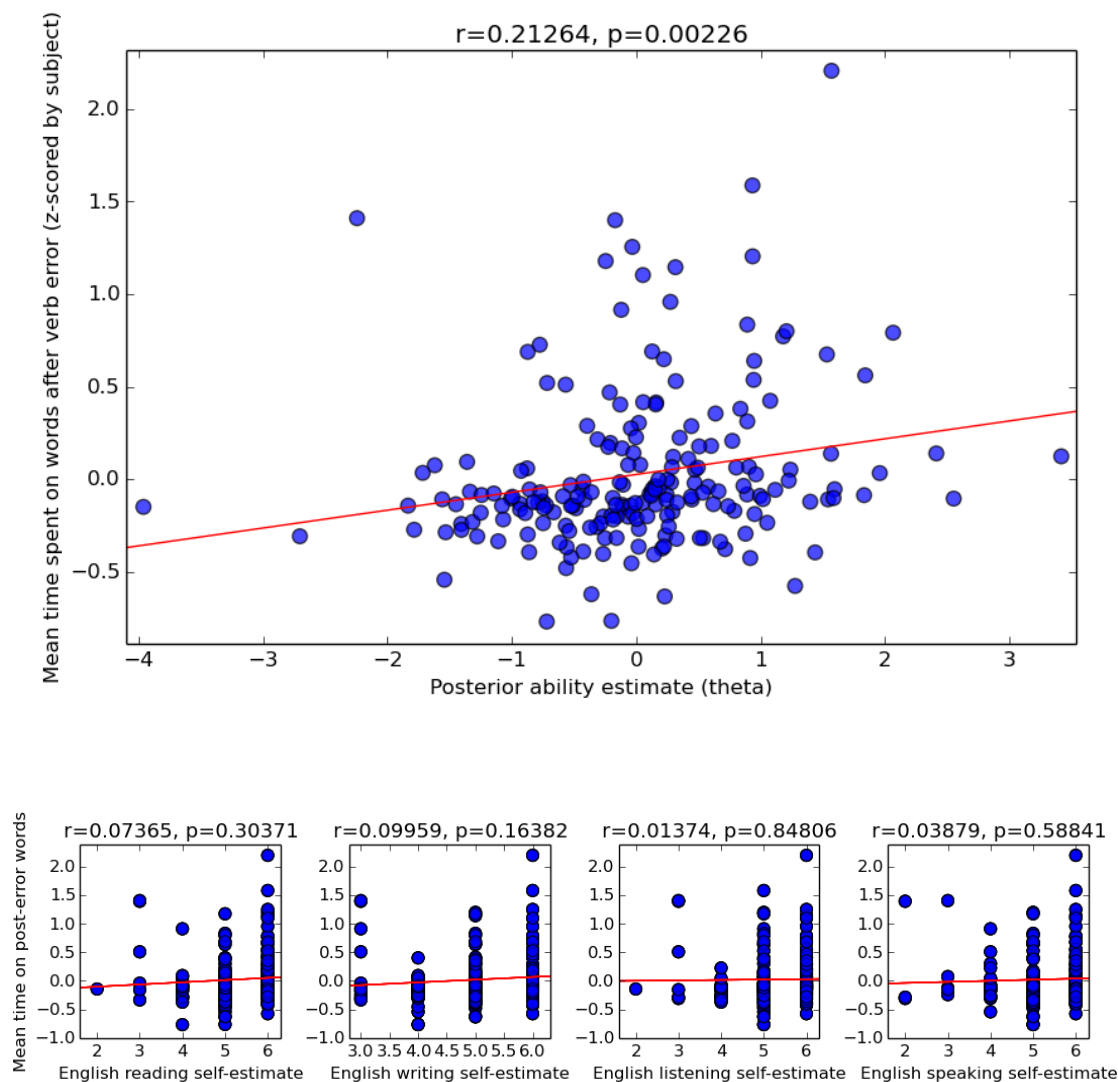


Figure 33: Scatterplots and correlations between self-paced reading syntactic error response times, and our posterior estimated ability (upper) and self-reported proficiencies (lower)

ESL Genie conclusion

In this chapter, we explored methods for adaptively assessing ESL learner knowledge, towards the goal of developing an Intelligent Tutor system with educational interventions to correct persistent grammatical errors. In the process,

we designed a system for smoothly quizzing a user, achieving competing design objectives such as simultaneously maximizing both throughput and user experience. A question bank that we developed on the basis of errors identified in a large corpus of transcribed ESL essays was used to collect data from UCSD undergraduates with an ESL background. Through leveraging Bayesian Item Response Theory, we found an efficient question-selection algorithm to minimize the number of questions needing to be asked in order to reach a confident estimate of a learner's ability. Finally, we found a promising relationship between subjects' predicted ability and their slowdown in response to an error in a self-paced reading task, suggesting the potential for self-paced reading to be used as a low-stress diagnostic instrument within an adaptive language learning framework.

The work described up to this point has been motivated by the goal of enabling quality learning opportunities for those who, for financial or other reasons, might not otherwise have access to high-quality tutoring, through establishing an open-source, freely accessible, adaptive tutoring system available over the Internet. The remainder of this dissertation examines the design challenges arising from the issue of limited global Internet access, and explores solutions that can overcome these barriers to bring the types of learning tools described here to the populations they could most benefit.

Chapter 3: Online education and the digital divide

Open Educational Resources

The Internet has fostered a wide community of resource sharing, and it is not difficult to find “free” content on just about any topic and in any media. Unfortunately, regardless of a content creator’s underlying intentions, by default any content he or she uploads and posts to the web is covered by implicit “all rights reserved” copyright (Smith, 2013), and cannot be redistributed, remixed, or broadcast without the creator/copyright holder’s explicit permission. While for the average content consumer, this may not make much practical difference, it puts a damper on efforts to repurpose and remix content to create new works, as well as for the open distribution of the content, which is of particular relevance to us as we examine ways to distribute educational content through offline channels -- how can someone request permissions for adapting for further distributing a work, when they are inherently disconnected from the content creator?

Creative Commons, established in 2001, released the first set of Creative Commons (CC) licenses in 2002 (“History - Creative Commons,” n.d.). These licenses allow a content creator to release content for free and open distribution, with “some rights reserved”, specifically choosing whether or not to allow the content to be modified, shared without attribution, used commercially, or released under a different variant of the license after modification (“About The Licenses - Creative Commons,” n.d.).

Around the same time that Creative Commons was getting started, MIT made the decision to release all of their course materials openly on the web, in what became MIT OpenCourseWare (OCW), “opening a new door to the powerful, democratizing, and transforming power of education” (Vest, 2001). Together, these two initiatives provided the momentum that would soon grow into the Open Educational Resource movement (Smith, 2013).

You say you want a revolution: Khan Academy, and the MOOCs

The online learning revolution

There has been a lot of excitement around the “online learning revolution” that has been unfolding over the past few years, and one of the primary virtues being celebrated is the potential for ubiquitous access. Khan Academy, for instance, is founded on the mission of “providing a free world-class education for anyone anywhere”, and Coursera aims “to give everyone access to the world-class education that has so far been available only to a select few”. An emphasis on universality is critical for any strategy that hopes to use education to level the global playing field, as we otherwise run the risk of perpetuating the digital divide by inadvertently but specifically neglecting those disadvantaged and isolated populations that could benefit most from access to these educational resources.

Pedagogical explorations: self-paced learning and flipping the classroom

“It does not matter how slowly you go as long as you do not stop.”
– Confucius

Self-paced, or mastery-based, learning involves allowing students to move through a curriculum at their own pace, spending more time on materials they are having trouble on, or moving ahead to new topics even if the rest of the class is not yet ready, or if the topic is a digression from the main curricular track.

Approaches such as those we discussed in Chapter 2 can be of great use for this, in efficiently determining where a user’s current knowledge fits within a domain. Tools and visualizations for allowing a coach (and the learner her or himself) to introspect the student data being collected by the system can play an important role in guiding the mentoring process.

Flipping the classroom, a concept that was described around the turn of the century (Lage, Platt, & Treglia, 2000) but that Khan Academy helped to popularize in recent years, involves inverting the traditional classroom model. Rather than class time being spent on teacher-centered, lecture-based activities, and then time after school spent doing exercises and other homework, the flipped classroom model has the lecture-style content in a format that can be viewed by students at home, on their own time, freeing up class time for supervised exercises, peer activities, and one-on-one mentoring. This model tends to rely heavily on educational technology, both for watching videos at home, and to allow for student progress to be tracked dynamically enough that

the teacher can act as an effective coach in the classroom, even for larger classes.

The digital divide

The historical divide

Upon visiting the New York World's Fair of 1964, science fiction visionary Isaac Asimov was inspired by the technological progress he saw displayed from the first half of the 20th century, but dismayed that they had not included predictions of where all this technology was leading, and what the world might look like 50 years down the road. Along with his own prescient foretelling of such things as video conferencing, 3D flatscreen TVs, and "mock-turkey", Asimov also offered a stark reminder that technology does not tend to evenly distribute its benefits:

Although technology will still keep up with population through 2014, it will be only through a supreme effort and with but partial success. Not all the world's population will enjoy the gadgety world of the future to the full. A larger portion than today will be deprived and although they may be better off, materially, than today, they will be further behind when compared with the advanced portions of the world. They will have moved backward, relatively. (Asimov, 1964)

The spread of television during this period serves as a useful example of the point Asimov was making. At the time Asimov wrote this, television penetration was rapidly rising in the United States, and by 1969, 97% of American households had a television, compared with 71.8% in 1956 (van Zuilen, 1977). In contrast, the first television broadcast in the African continent took place in

Nigeria in 1959 (Uche, 1989), and still today, television penetration through the developing world is at approximately 73% (Looms, 2011). Given the role that television historically played in supporting the spread of literacy, education, and news, the 50+ year lag in distribution serves as a clear case study of the often glacial pace at which technology trickles down to the less advantaged.

The modern instantiation of the digital divide revolves largely around access to, and ability to take advantage of, the Internet. The following sections will outline the origin and design that underlies Internet infrastructure, how the Internet has developed since its inception, and the types of access disparities we observe across geographic and socioeconomic groups.

The days of proto-Internet

The first computers that began to emerge in the 1950s and 1960s were room-sized, immovable monoliths, isolated from one another, and at first usable by only one person at a time, as depicted in Figure 34(a) below. Programs were run in “batch processing” mode, punched into paper cards and physically transported to the computing center for processing one after another. Later, from the 1960s into the 1970s, the development of techniques for task-switching and time-sharing allowed multiple users to simultaneously connect to a mainframe computer via “dumb terminals”, devices that typically had a screen and keyboard, but very limited processing power, as shown in Figure 34(b). Typically, these terminals were located somewhere near the mainframe computer they were connected to, for example on the same campus, although there were some early

experiments in longer-distance terminal connections over slow “teletypewriter” lines (Cooper & Latham, 1967; Marill & Roberts, 1966).

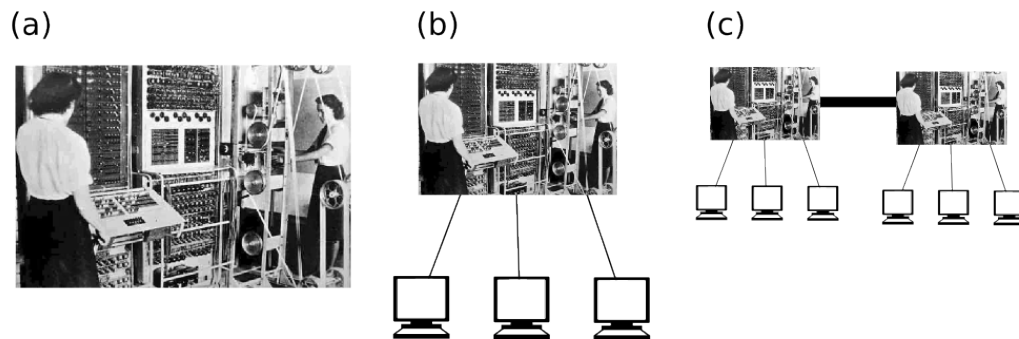


Figure 34: Mainframe topologies

One of the challenges faced during this period was that the computer hardware and software configurations being used on campuses and by businesses were highly heterogeneous, and lacked the shared standards that would have allowed for easy reuse of programs across facilities. The inefficiencies that resulted from this need to perpetually reinvent the wheel served as one of the early driving forces behind the development of networks that could connect multiple mainframe computers together, as seen in Figure 34(c), and some envisioned a world where a user could log into their local mainframe, and from there access services and data from other mainframes via a globally interconnected network (Marill & Roberts, 1966). Some of the first uses of these low-bandwidth connections were for educational purposes, with Europe’s “first home computer terminal” being used by a child in the family to

practice math sums, inspiring his father to imagine the role such systems might play in households of the future:

Rex Malik sees a future world where children could be virtually educated by computer, where every home will have its own terminal, plugged into a central brain, and from the brain will come not only school lessons; he sees his son growing up in a world where eventually his very thoughts could be stored and perhaps assessed for his future use. (Cooper & Latham, 1967)

The emergence of the Internet

A pivotal point in the transition from these early point-to-point networks to what we now know as the Internet came with the development of packet-switching networks, which allowed data to be broken up into chunks and routed indirectly from one computer to another via hops through other intermediate nodes. Baran (1964) described various network topologies, and conducted simulations to demonstrate the effectiveness and resilience of local, decentralized routing algorithms, even in the face of unreliable or changing communication links, or the destruction of a subset of the nodes in the graph.

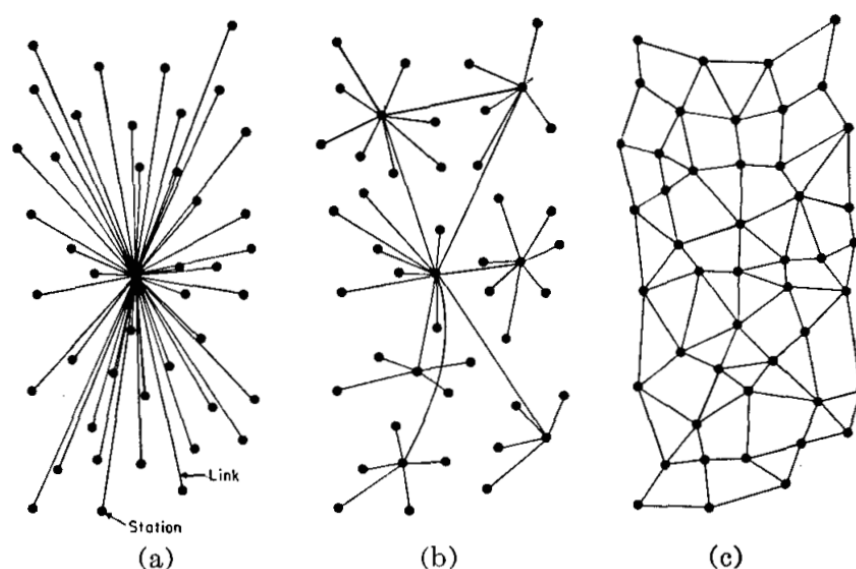


Figure 35: Network topologies in the continuum from (a) centralized networks, to (b) decentralized networks, to (c) distributed networks (from Baran, 1964)

Access to the Internet

Technology is not only a parent of wealth and development (creates it), but also its child (stems from it). It has therefore the potential to spawn either a virtuous circle or a vicious circle between existing or missing development and technology. [...] Either the excluded could be armed with new empowering tools despite their unfavorable starting position or their relative situation could worsen while the benefits once more enable the already well-off to make headway. (Hilbert, 2010a)

One of the most challenging barriers for enabling equitable access to online learning resources is the massive disparities in Internet connectivity between developed and developing regions, with approximately 65% of the world

still not classified as Internet users, according to the World Development Indicators compiled by the World Bank.⁹

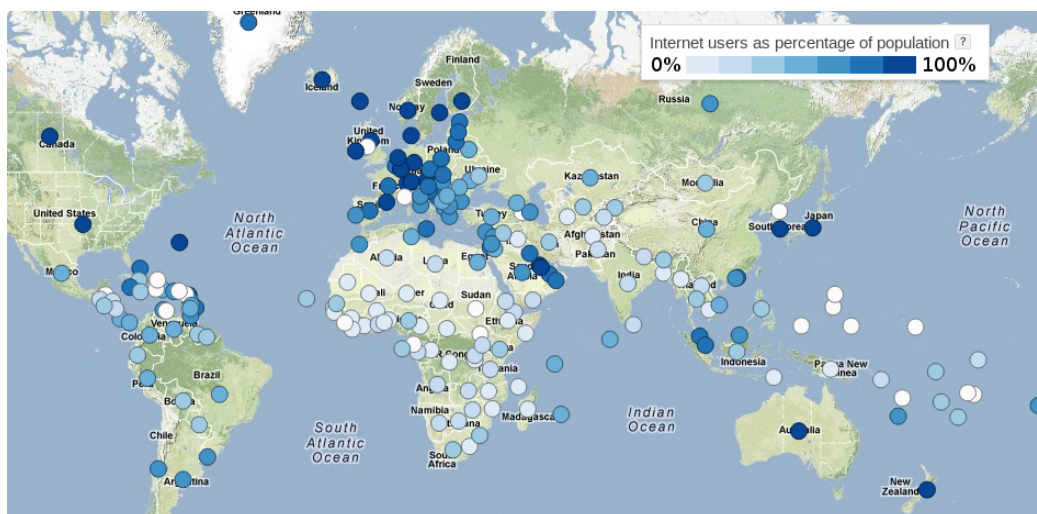


Figure 36: Global distribution of Internet users by country as a percentage of population

The countries with the least Internet access are the same countries likely to have fewer educational resources, such as teachers and books, available to learners. This can be poignantly visualized by comparing countries' Internet penetration rates with their pupil/teacher ratios (see Figure 37 below), and noting that the countries with the fewest teachers per student also have the least Internet, and hence the populations that could potentially most benefit from alternative educational resources are the same ones that are least able to access them.

⁹ <http://data.worldbank.org/data-catalog/world-development-indicators> (graph generated using Google Public Data Explorer: https://www.google.com/publicdata/explore?ds=d5bncppjof8f9_)

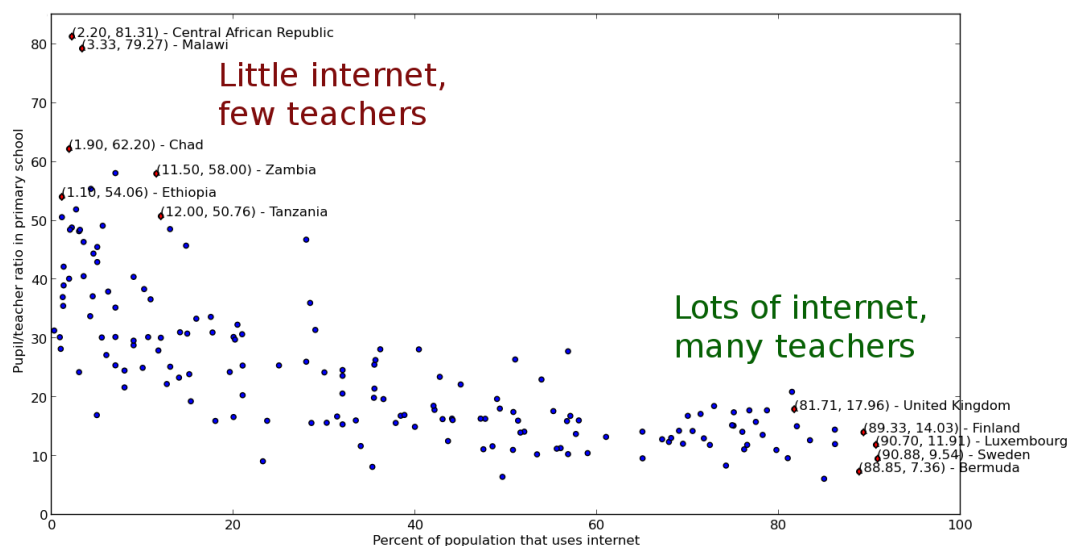


Figure 37: Scatterplot showing the relationship between student-teacher ratios and Internet connectivity in a country

If we're serious about achieving the online learning revolution's mission of universal access, and don't want to leave further generations of students behind by waiting for high-speed Internet to reach everyone, then we need to explore solutions for distributing and hosting open educational resources via low-bandwidth and offline channels, taking advantage of low-cost or pre-existing infrastructure.

Where is the "cloud"?

In recent years, "moving to the cloud" has become a popular buzz phrase, particularly within business and technology circles, but also as a household expression in many areas. When an Internet service is "in the cloud", it means it is being hosted in a large datacenter with shared infrastructure, allowing for

economies of scale to kick in as demand for one service fluctuates, at which point it can allocate more continuously billed resources for itself to cover the spike, before releasing them for others. The downside of centralizing web services in this way is that they tend to be focused around the US or Europe, leading to long-distance routing and slower connections for those in less developed areas. There are also political and privacy concerns about having so many of the Internet's core data centers located such that data passes through countries that may be monitoring or tampering with it (Kravets, 2013).



Figure 38: Location of Google's data centers¹⁰

From the map of Google's data centers shown above, it is clear that large swaths of the globe are located far away from any of these datacenters, and hence will experience much slower access times and potential outages.

¹⁰ Image source: <http://www.google.com/about/datacenters/inside/locations/>

Facebook, similarly, only very recently opened its first datacenter outside the US (Lloyd, 2013).

As a further test, I ran an IP address lookup for the top 100 most popular websites, from within a server physically located in India, to see where requests to these websites would be routed when initiated from South Asia. Of the 83 resulting addresses for which I was able to successfully locate the home country, almost two thirds of them were centered in the United States (see Figure 39, below).

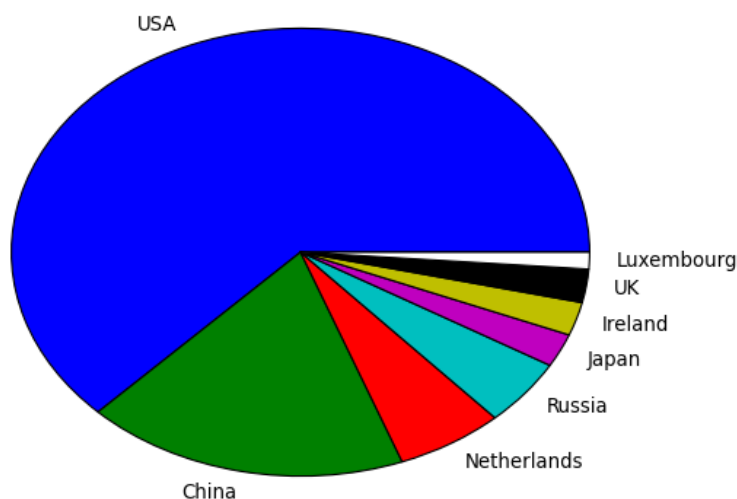


Figure 39: Geographic distribution of servers hosting the 100 most popular websites, when accessed from a computer in India

Despite the popularity of “the cloud”, and the centralization that it brings along with it, there is a small but growing movement to, as one group phrases it, “redentralize” the Internet, by building technologies that support running websites and services that are frequently located in the cloud (such as email,

social networks, and blogs) from a local server. Two of the top motivations that have promoted this resurgence of interest in decentralization are: 1) putting control of data back into the hands of users (of particular interest following the surge in privacy concerns caused by recent revelations about the scope of data monitoring by government agencies), and 2) allowing for connectivity that is resilient in the face of interference such as natural disasters or censorship.

Acknowledging the digital divide

The severity of the digital divide is often downplayed by those who are reassured by the growth trends they see in the developed world, and either assume that the same progress is being made, at perhaps a somewhat slower rate, in the developing world, or, under a somewhat less charitable interpretation, are not concerning themselves with inequalities that lie outside the scope of their home countries. In a study by DiMaggio and Hargittai in 2001, for instance, they argue that we should start to focus less on the issue of *access*, and more on the differing ways in which those people *with* access are making use of it (something they refer to as “digital inequality”), given that “access to and use of the Internet has spread widely and swiftly” (DiMaggio & Hargittai, 2001).

At that point, access to the Internet could no longer contribute significantly to social inequality, simply because nearly everyone would have it. Like efforts to extend telephone service, the attempt to ensure that every American could go on-line, while important from a policy perspective, would represent a mopping-up operation.

But would this mean that the “digital divide” had been overcome, in the sense that equality of access to the benefits of the Internet would have been achieved? Some policy analysts have implied that

this is the case. Drawing on the history of telephone access, (Compaine, 2001) argues against legislation to ensure universal access because the combination of market forces and government programs currently in place are achieving that goal already.

While they go on to make some good points about the inequalities that can exist within groups that do already have access (such as differences in device quality, context of use, purpose, and skills), their exclusive citation of American-centric data, and subtle oversight of the the global realities of the digital divide, can be highly toxic in terms of shifting the dialogue away from the highest variance inequalities, and focusing on nuances that, while important, may often just lead to a shuffling of the upper stack of cards in the deck.

The rapid spread of the Internet through the developed world does indeed pose challenges for those of us living in these privileged regions to fully comprehend or accept the scope and severity of the digital divide's effect on global inequality. I have heard people estimate that the Internet should be fully ubiquitous within 10 years, but neither the history (e.g. the slow penetration of television into poorer areas around the world, with rates in the developing world only *now* reaching the levels seen in the United States in 1956), nor the quantitative data about Internet growth rates, should be leading us to such a rosy estimate of the timeline for universal Internet.

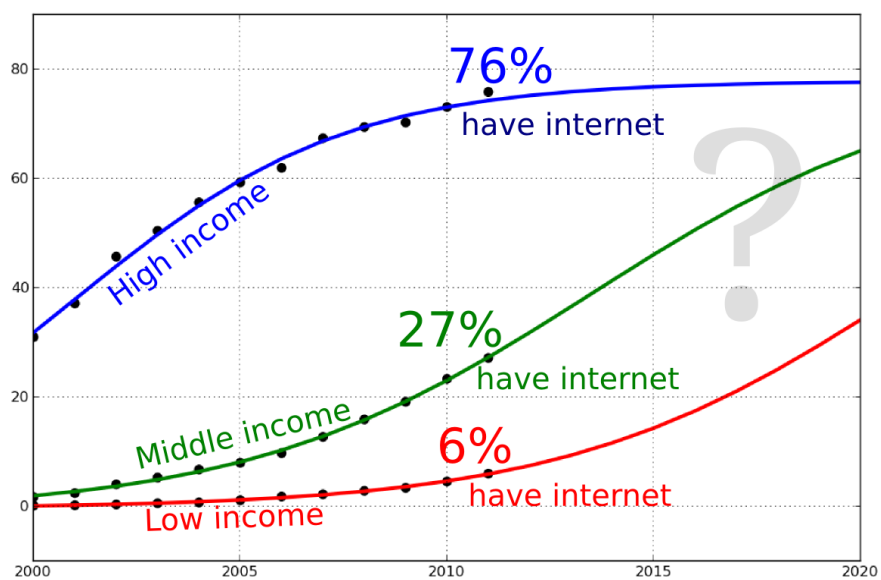


Figure 40: Internet penetration levels in low-income, middle-income, and high-income countries¹¹

¹¹ Data from the World Bank's WDI dataset
(<http://data.worldbank.org/data-catalog/world-development-indicators>)

Chapter 4: Taking the online learning revolution offline with KA Lite

Identifying opportunities and approaches

On the one side of the equation, we've seen that 2/3 of the world is lacking Internet, and facing critical educational challenges, while on the other side, we've seen the fantastic progress that groups like Khan Academy are making in developing high-quality Open Educational Resources and releasing them online. The broken part of this equation is that those who are in the most dire need of educational opportunities are not able to take advantage of these new materials and tools. What this means is that, ironically, the online learning revolution ends up primarily benefiting those already in the upper ranges of global privilege, despite the firm intention on the part of the organizations creating and sharing this content to reach these underserved populations.

What then can we do? Several options are on the table. The first, and most commonly chosen by default, is simply to wait until the Internet's infrastructure has grown to the point where it truly has reached some level of ubiquity. As we discussed previously, estimates of 10 years are likely to be overly optimistic, and even then, we'll be leaving generations of learners behind in the meantime, forcing them to play catchup along whatever the next big dimension of differentiation will be by that time. The second option is to be more proactive in extending higher bandwidth Internet to reach those who currently don't have access. This is the approach being pushed by several groups, particularly the big players in the online advertising space, such as Google (with Google Fiber,

domestically, and Google Loon, internationally) and Facebook (via Internet.org, a consortium they formed with several mobile phone companies). Mark Zuckerberg (2013), in a whitepaper entitled “Is Connectivity A Human Right?”, outlines some of the approaches Facebook hopes to take in increasing global connectivity.

While initiatives like these are an important part of pushing global infrastructure forward, we saw in Chapter 1 that calling something a human right does not magically teleport us into a world where that goal has been met, so we need to be looking to other interim solutions that can be accomplished with less upfront requirement for large capital investment and political maneuvering. The third option, then, is the one we are pursuing here: leveraging Open Educational Resources (described in Chapter 3) combined with the rapidly falling prices of low-cost hardware devices (described in the next section) to rapidly bring opportunities to engage with high quality educational materials and tools to those for whom Internet connectivity is unavailable or unaffordable. This chapter will describe the work we have done in designing, developing, and deploying KA Lite, an offline cross-platform software solution to enable local access to Khan Academy’s collection of thousands of educational videos and exercises.

The rise of low-cost computing

The parallel revolution

In Chapter 3, we examined how the increase of widespread broadband Internet access throughout the developed world has helped to fuel the surge in

popularity of online learning resources that we are seeing today. While in most of the world, Internet connectivity remains slow, spotty, and expensive, a parallel technological development has important implications for our ability to bring educational resources to the disconnected world: the prices of ever-more-powerful mobile devices, particularly Android phones and tablets, have been continuously plummeting in recent years, with a corresponding upsurge in adoption.

In the following sections, I will provide some background on the low-cost computing movement, by examining several of the devices that have been designed with low cost as a primary driving factor, and the role they have played in efforts to bring educational technologies to disadvantaged communities.

One Laptop Per Child

The One Laptop Per Child (OLPC) project was first announced in 2005, with the goal of producing a rugged, versatile, low-power, educational computer aimed at children in the developing world. When it was announced, the price of the device was planned to “start at approximately \$100 and then steadily decrease” (2005), which at the time was impressively ambitious, and fueled a media frenzy that helped propel the non-profit coalition behind the project into the spotlight. However, the \$100 price point turned out to be *overly* ambitious, and in the 8 intervening years since it was announced, the price of the OLPC has yet to drop below \$200USD (Alvarez, 2013). However, despite its failure to reach its target price, the OLPC and the attention it attracted helped to foster discussions

about the importance of low-cost devices, and the goal of making computing accessible to disadvantaged populations.

Android and the Aakash

Android is an open-source operating system for mobile devices, such as phones and tablets, developed and supported by Google along with a consortium of device manufacturers and mobile service providers. The Android OS was first announced in 2007 (Open Handset Alliance, 2007), and the first phone running Android came out in 2008. In contrast with Apple's unified hardware strategy for the iPhone (running the proprietary iOS platform), which was first released in 2007, the Android ecosystem was explicitly established so as to foster a multi-manufacturer competitive marketplace, which quickly served to drive down prices and increase the diversity of devices being made available. In 2010, Android began a meteoric rise in global popularity, quickly overtaking the previous leader, Symbian OS, to become the most popular smartphone OS around the world (see figure below). As of late 2013, Android now commands 80% of global smartphone market share (Harper, 2013).

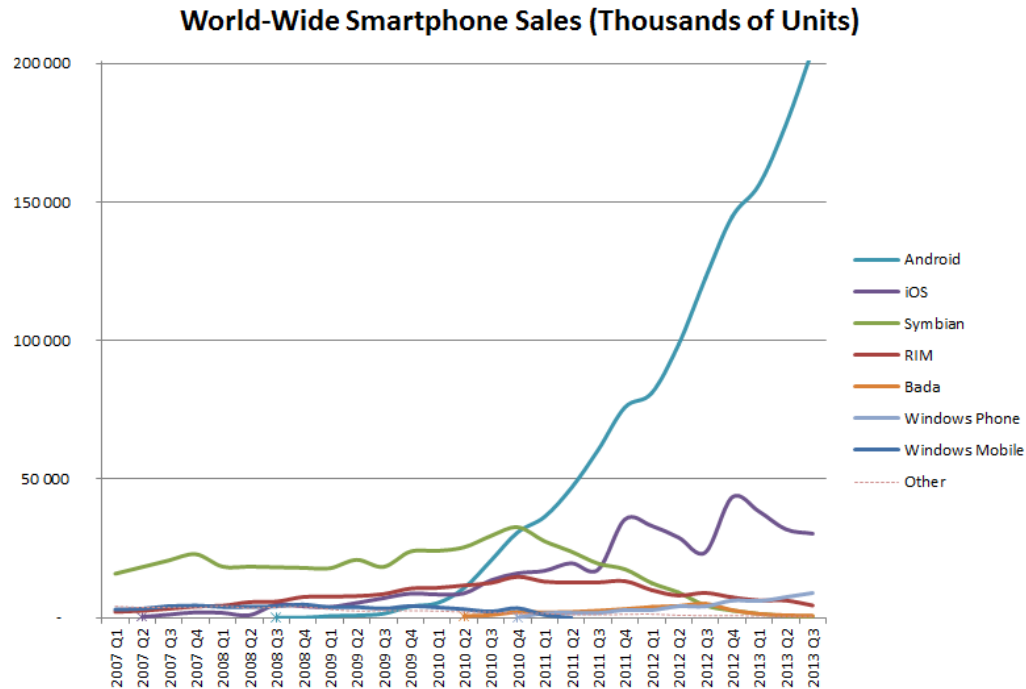


Figure 41: Worldwide Smartphone Sales, from "Gartner Smart Phone Marketshare 2007-2013"¹²

Much of this popularity comes from growth in the developing world. While Apple's iOS still dominates in the developed world -- particularly North America, Western Europe, and Australia -- the rest of the world is rapidly being taken over by devices running Android.

¹² Compiled at http://commons.wikimedia.org/wiki/File:World_Wide_Smartphone_Sales.png

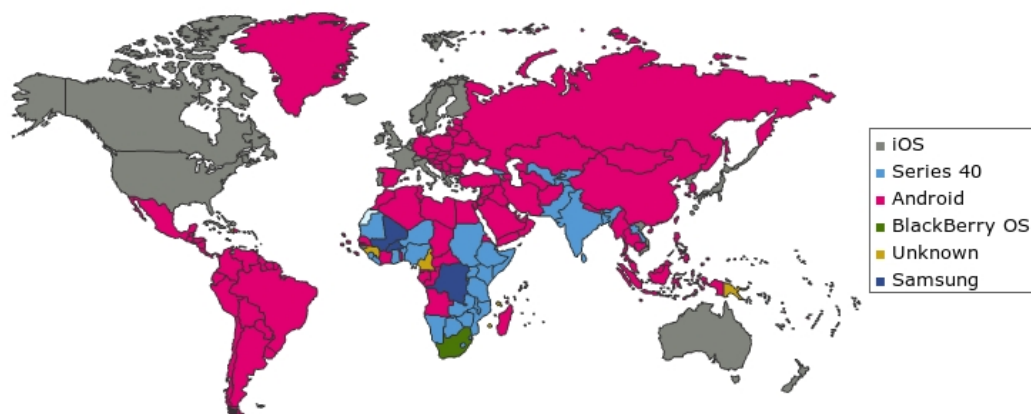


Figure 42: StatCounter Global Stats: Top Mobile & Tablet Operating Systems Per Country, Nov 2013¹³

A large part of the growth seen in the global Android market can be attributed to declining costs; the average selling price of Android devices is dropping by 8-9% per year (Miller, 2013), with the market now being flooded with tablets in the \$50-\$70 range. One tablet in particular, the Aakash, has been pushing these costs even lower with the support of the Indian government, with a target price of \$20 in India. An international version of the tablet, built by Ubislate, can now be purchased in the United States for \$38 (Chicago Tribune Company, 2013).

¹³ <http://gs.statcounter.com/#mobile+tablet-os-ww-monthly-201311-201311-map>

The Raspberry Pi

This credit-card sized computer was the original inspiration behind KA Lite. Costing \$25/\$35 (model A/B), the Raspberry Pi boots from an SD card into a graphical Linux-based operating system geared towards education, including software such as MIT's Scratch¹⁴, and has connectors for HDMI (to play high definition video to a projector or screen), composite video (to output to a TV), an audio jack, ethernet, and USB ports. Running on as little as 1 Watt of power, the Pi is a good match for electricity-constrained environments. The Pi, designed and produced by a UK charity, the Raspberry Pi Foundation, was envisioned as a \$25/\$35 from the start, and although it took several years to go from prototype to production, they stuck to this ambitious goal and delivered the first Pi's in early 2012. Since then, they have sold over two million of the diminutive devices, all around the world (Upton, 2013), where they're being used as the brains of robotics projects, as cheap desktop computers by being plugged into old TVs, to take photos from space balloons -- and as servers for KA Lite!

¹⁴ <http://scratch.mit.edu/>

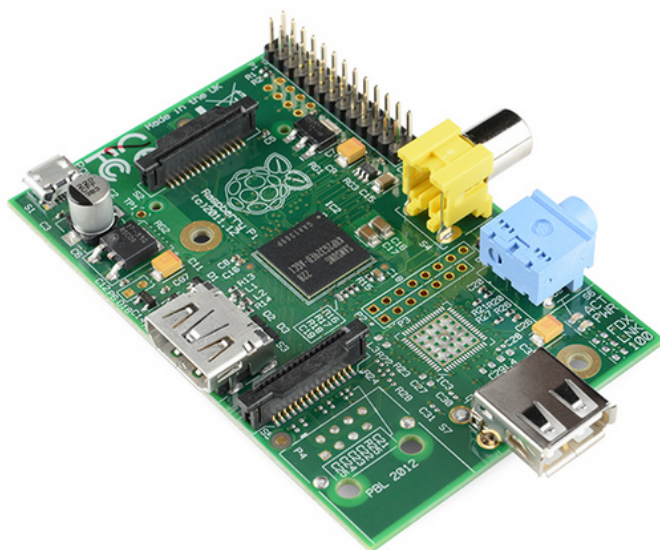


Figure 43: Raspberry Pi Model A board¹⁵

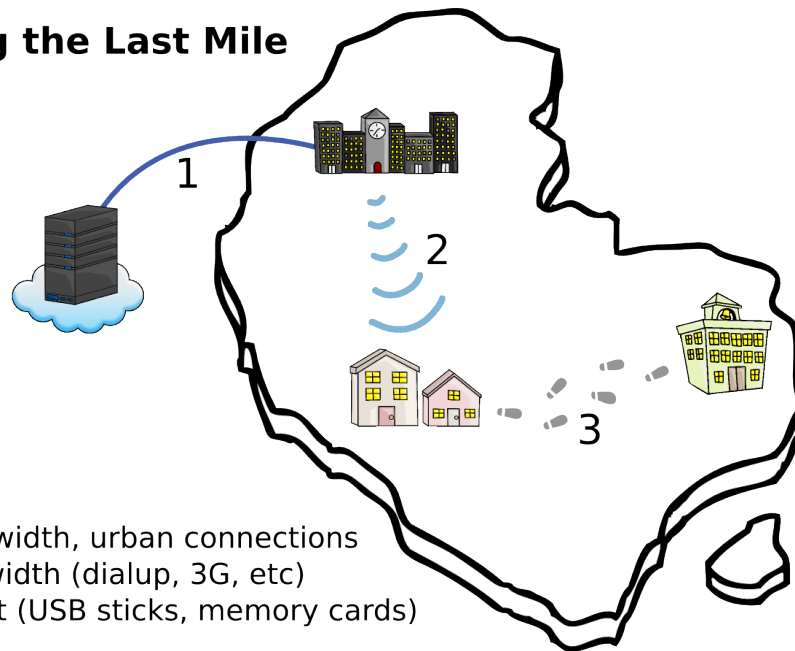
Design considerations for KA Lite

Be able to install and work offline

Clearly, one primary need of a system like KA Lite is that it be able to fully function when there is no access to the Internet at all. For this to be possible, we have to have a way to get the software and content to its final destination. Luckily, in most cases, we are dealing either 1) with simply a low-bandwidth situation, where content can be slowly downloaded once and then used repeatedly from a local server, or 2) with a location that is reachable by physical transport, whether it be walking, vehicle, or even, as Sal suggests in his book, by donkey (Khan, 2012).

¹⁵ Photo credit: SparkFun Electronics, <http://flickr.com/photos/41898857@N04/8747111043>

Reaching the Last Mile



1. High bandwidth, urban connections
2. Low bandwidth (dialup, 3G, etc)
3. Sneakernet (USB sticks, memory cards)

Figure 44: Modes of delivering software/content to the final destination

Basic features like video watching and exercise interaction

The KA Lite web platform was designed to include many of the core features of the Khan Academy website, in a simplified form. Students can navigate through the content either by the video topic tree or the exercise knowledge map, and jump between related videos and exercises to move back and forth between “practicing” and “reviewing”.

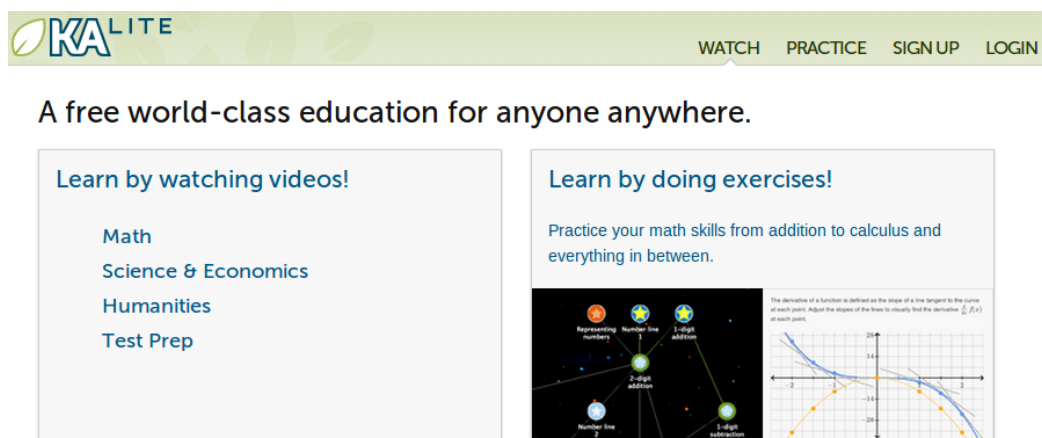


Figure 45: KA Lite homepage, showing entry points into the content

Progress tracking, points, and coach reports

Beyond the core video and exercise content itself, one of the highest value components of a system like Khan Academy -- and one that is necessary in order to properly support pedagogical innovations such as self-paced learning and flipping the classroom -- is the ability for learners to have their progress tracked as they navigate through the curriculum, and then be able to visualize their current state. Additionally, it is critical to provide “coach report” tools that enable teachers, mentors, or parents to visualize both individual-level and aggregate student progress data, as these tools allow them to quickly and easily assess what their students are struggling with the most, and intervene to offer guidance.

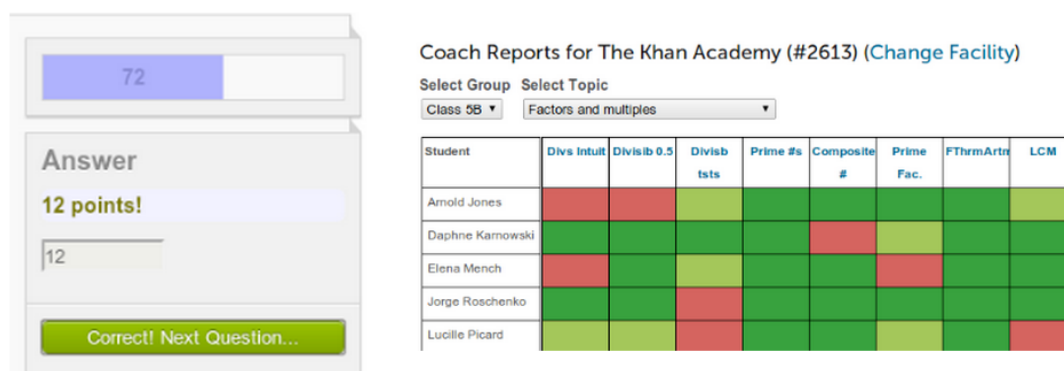


Figure 46: KA Lite point system and coach reports

Support content updates

Administrators can choose which video content (and subtitles) to load onto the local server, if there is occasional Internet connectivity (otherwise, they can copy these in from a USB stick that was carried in).



Figure 47: Interface for downloading videos and subtitles (left) and portion of video showing Chinese subtitles (right)

Data synchronization (eventual consistency)

We want to include mechanisms for peer-to-peer synchronization of software updates, content, and usage data. In offline deployments, there is a risk

of a) the software languishing due to a lack of update channels, and b) not having a centrally measurable impact due to siloing of usage data. It is critical to have low-friction ways to effectively leverage offline/sneakernet channels -- such as USB sticks, local mesh networks, and roving servers -- to keep devices in sync and link them back up to the Internet.

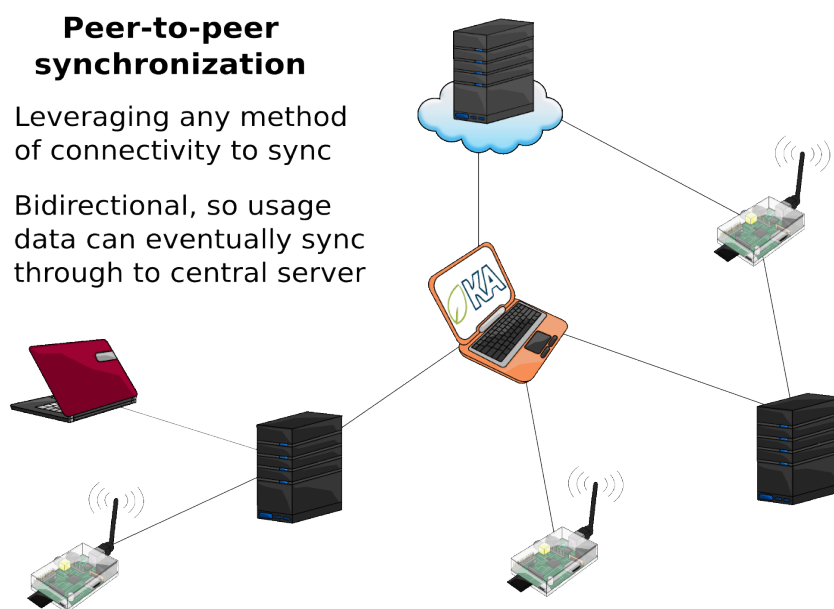


Figure 48: Peer to peer synchronization network

The KA Lite platform is currently able to take advantage of low-bandwidth Internet connections to trickle-synchronize content and usage data with our central server, but a future item on our development roadmap is the ability to leverage local area network connections and removable storage media such as USB sticks, to synchronize in a fully peer-to-peer fashion, administered entirely through the local KA Lite web interface.

Cross-platform support (portability)

Be able to run on any computer operating system. With the diversity of operating systems and versions that are in use through the developing world, and the difficulty of upgrading system software in offline contexts, broad OS support is critical.

In order to fulfill this goal, we made the decision to restrict all of the server-side code to pure Python, with no dependencies on other software and no platform-specific compiled modules. This reduces the friction of installation and deployment by not requiring that the software that is part of a typical web stack to be installed/configured, such as Apache/modwsgi, MySQL, and a cronserver. Instead, we have integrated pure-Python solutions for all of these components, which means that any system with Python installed can directly run the server software.

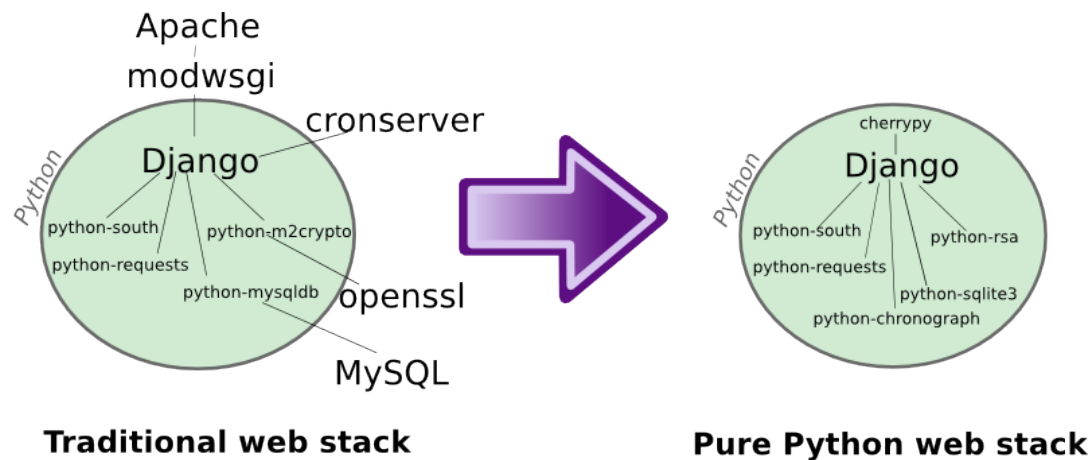


Figure 49: Our simplification of the traditional web stack to make it into a pure Python codebase that can be installed and distributed easily

Running on low-powered and inexpensive hardware

In many areas of the developing world, especially the offline communities being served, electricity is either intermittent or limited, so the software should run on hardware that can run effectively on battery, solar, or generator power.

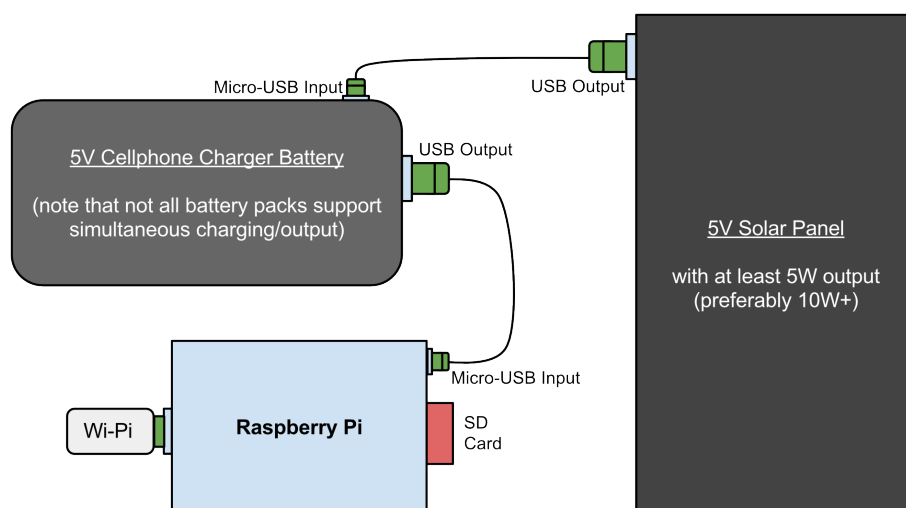


Figure 50: A hardware configuration using the Raspberry Pi as a local Wi-Fi access point, powered by a solar-charged cellphone charger battery

In order to reduce costs and thus be able to reach as many people as possible at the lower rungs of the socioeconomic ladder, we want to be able to run on inexpensive, easily sourced or existing hardware (such as an old Windows desktop in a computer lab, or a donated Mac laptop). For deployments in areas without existing infrastructure, low-cost solutions must be available and supported, to maximize the impact of any resources invested in hardware.

Ease of installation process

Since many of the end-users who will be installing the platform will not have had experience installing or configuring server software, the installation process needs to be smooth and automatic, through a graphical interface. We are part of the way towards this goal, with a simple graphical installer for Windows, but aim to make cross-platform installers and updaters in the near future.

Internationalization

We need to be able to support internationalized interfaces and content, especially as Khan Academy volunteers and partners have translated Khan Academy content into so many languages. The diverse linguistic backgrounds of offline communities clearly necessitate a solution for the inclusion of translated content and multilingual interfaces. The internationalized version of KA Lite is now in the process of being released.



Figure 51: KA Lite main page translated into Bulgarian

KA Lite deployment contexts

Classrooms: Burmese monastery school

At a monastery school in Burma, donated laptops are being used to engage with KA Lite, along with translated videos coming out of the Khan Academy Burmese Translation Project. With only 5 laptops shared amongst the students at the school, learning naturally becomes a highly social activity, encouraging discussion, questions, and collaboration.



Figure 52: Students in a Burmese monastery school using KA Lite¹⁶

¹⁶ Photo credit: Ashley Takami

Classrooms: Nalanda Project

“You might be poor, your shoes might be broken, but your mind is a palace.” — Frank McCourt

In India, in an initiative dubbed the Nalanda Project, we are partnering with the a family foundation called Motivation for Excellence (MFE), and the educational non-profit organization Akanksha Foundation, to deploy KA Lite to Grade 4 and 5 classrooms in Akanksha’s network of schools serving children from low-income communities. These classrooms are equipped with a Raspberry Pi computer set up as a local wifi hotspot, running KA Lite. Each student uses an Aakash tablet to connect to the Pi’s wifi hotspot, and access the locally running KA Lite server through their tablets’ browsers (with no Internet required or involved in the classroom at any point). The total one-time hardware cost for a configuration like this is around \$2,500 USD (sponsored by MFE), but with the majority of that cost being from the 35 tablets, and the prices of these tablets dropping rapidly, this cost should soon be quite a bit lower. Additionally, with each hardware set being shared amongst several classrooms, and the costs spread out over hopefully at least 2 years, the cost per student can hopefully be brought below the magical amortized price of \$10 per student per year, at which point it has been said that communication technologies can shift from being luxury items to being necessities (Hilbert, 2010b).

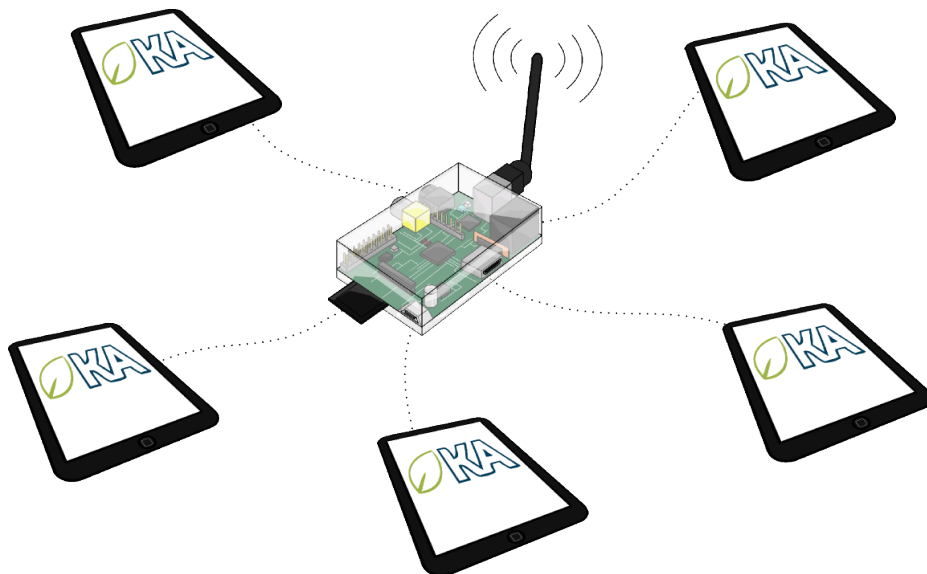


Figure 53: Hardware configuration used in Nalanda Project classrooms, with 1 Raspberry Pi wifi access point server running KA Lite, and 35 Aakash tablets, one for each student (5 shown)

While the servers in these deployments have been preloaded with the majority of the Khan Academy content, the classes have been focusing on the math content, in particular, as this is the curriculum for which Khan Academy has the most coverage, especially in terms of interactive exercises. The teachers in the first three classrooms each implemented KA Lite slightly different with their students, with one focusing on the videos (and having, for example, students watching and trying to shout out the answers to sums before the video finished), and another putting more emphasis on the exercises, having students refer back to the videos as needed when they didn't understand a concept. Teachers also supplemented time spent with KA Lite with some of their more traditional lecturing, though they were now able to use the coach reporting systems to

determine what students were struggling with overall, and cater to their needs in a data-driven fashion.



Figure 54: Students in an Akanksha school in India, using KA Lite on Aakash tablets¹⁷

Students seem to be highly motivated and engaged with KA Lite, Vivek Ragavan, from the Motivation for Excellence foundation, had the following reaction after visiting the classrooms recently:

“I was absolutely amazed at the overwhelmingly positive response by the kids and teachers to the KA Lite platform and Aakash tablets. All the kids loved the tablets and systems, were greatly appreciative of having the opportunity to use them. The level of engagement and motivation (especially with the points based mastery system) were

¹⁷ Photo credit: Vivek Ragavan

stunning. All the kids wanted to have the tablets for the rest of their school years!”

In a video from one of the classrooms, a student describes his newfound love for math:

Student: “We used to be very tired with this [pointing at blackboard], but when we get this tablet, oh, then we are not very tired like that.”

Teacher: “Which is correct, because they’ve never been able to do math like this. Never! I don’t know the last time they just sat for 3 hours and just did math, non-stop!”

In fact, motivation seems to be so high that one of the concerning observations that came out of recent feedback from the classrooms was that some students were getting so engaged and focused on their learning through KA Lite on the tablets that they may have neglected to *move*, leading to ergonomic issues. One student “didn’t want to get up from the tablet, [...] focusing on that only”, and ended up with a crick in her neck that she noticed in the dance class that followed. Another student learned her own ergonomic lesson after spending a class period answering math exercises so intently that when the class was over, it took a moment to be able to extend her arm fully again, so the next class period with the tablets, she made sure to stretch her arms from time to time. Considering these observations, it may be worth looking into including ergonomic reminders interspersed with the videos and exercises, encouraging stretching, breaks, and good posture. Teacher training on basic ergonomic issues could also be helpful, so they can monitor and guide their students.

As these classrooms serve as an excellent environment to test KA Lite and better understand, firstly, the degree to which it is having a net positive

impact, and secondly, what the critical mechanisms of the system are that enable these positive outcomes, we are partnering with Motivation for Excellence, UCSD Economics professor Karthik Muralidharan, and his PhD student Sarojini Hirshleifer, to conduct Randomized Controlled Trials of the Nalanda project as it scales up to 40 classrooms in June of 2014.

Orphanage



Figure 55: Children using KA Lite on donated laptops at an orphanage in Cameroon¹⁸

UConn Empower, a student group from the University of Connecticut, traveled to an orphanage in Cameroon (run by “A Better World Organization”), and setup 14 donated laptops, installed with KA Lite, for the children to use. They

¹⁸ Photo credit: Lior Trestman

arranged it such that the older children could help mentor the younger children, and take care of the computer lab. What happened next was particularly exciting:

In talking with the Director recently, he has informed us that many locals, including a number of teachers, come daily to watch the videos and practice the exercises as well; what was meant to be a classroom for some kids is now a classroom for the community.
(Lior Trestman, personal correspondence, 7/30/13)

Correctional facilities

“He who opens a school door, closes a prison.” — Victor Hugo



Figure 56: Entrance to a prison we visited near Boise, Idaho¹⁹

When we began the KA Lite project, prisons had not crossed our minds as a potential site for deployments, but they have since become one of the most

¹⁹ Photo credit: Julie Oye-Johnson

important contexts. Educational programs in prisons provide inmates with opportunities to continue their educational progress, which for many of them was abruptly ended earlier in their lives. Prison education increases marketable skills, self-confidence, and provides a productive outlet within the confines of incarceration. The United States sadly imprisons more people per capita of any country in the world, with over 2.2 million people currently being held in prisons and jails around the country. Furthermore, of the 700K prisoners released every year, 40% of them will recidivate, and end up back in prison within a 3 year period. A recent thorough meta-analysis of 30 years of research into the effectiveness of correctional education programs found that recidivism rates dropped by 43% when such programs were in place, which also means that funding education in prisons made sense even from a purely economic perspective: for every \$1 spent on prison education, incarceration costs were reduced by \$5 over the three year period post-release (Davis, Bozick, Steele, Saunders, & Miles, 2013).

KA Lite has been deployed in prisons in several states, including across the state of Idaho, where it in its first pilot in the summer of 2013, a teacher had 20 out of 20 students pass their math GED (high school completion exam), something which he had never before observed (Kamenetz, 2013). When we visited in January of 2014, one thing that quickly became apparent was the particular appropriateness of self-paced learning in the prison environment, due to the obviously high variance in skills and backgrounds amongst prisoners, some never having made past grade school, and others with advanced degrees.

One design consideration that is perhaps quite unique to the prison deployment scenario is the existence of a strong need for data synchronization, while at the same time a strict prohibition against any form of Internet access. From discussions with teachers and administrators working in the prisons in Idaho, we learned that the need for data synchronization is twofold:

1. Offenders are frequently transferred between facilities, either due to shuffling that is done to balance bed occupancies, or as part of an offender's transition towards preparing for post-release re-entry. With over 15,000 transfers happening across the state of Idaho every year²⁰, this frequently leads to frustration for both teachers and students, who have to start from scratch without a clear picture of past progress, every time a transfer happens. Being able to seamlessly transfer a learner's tracking data from a KA Lite installation in one facility to an installation in another facility would be a real boon for ensuring the continuity of a learner's trajectory.
2. Educational programs in prisons are chronically underfunded, and straddle the line between the worlds of Corrections and of Education, not quite fitting into either. Processes for gathering assessment and impact data are very far behind, but can play an important role in being able to demonstrate value being produced, to keep state and private funding

²⁰ From personal correspondence with Julie Oye-Johnson, director of the education program at the Idaho Department of Correction.

sources from cutting budgets even further, and there is a lot of room for innovation in software and tools to facilitate this process.

One of the primary challenges, of course, is that communications in and out of the facilities are tightly controlled, and computers on the internal network cannot be connected in any way to the external world, either directly or indirectly. As one administrator framed the problem: while there is interest in finding ways to bring the benefits of connectivity into the correctional environment, the moment an offender found a way to send a message to a victim, the public backlash would be so strong that the whole system would be shut down, and so very few organizations are willing to consider that risk.

One approach to this problem would be to take advantage of the “sneakernet”, and sync data from the server in the prison onto a USB stick, which would then be carried to an external computer and connected to upload to the syncing server. The main issues with this method are logistical, with multiple trips in and out of an (often somewhat isolated) facility in order to sync data in both directions, and more manual work and room for error in carrying out the procedures.

An idea for an alternative approach came from visiting the prisons themselves, and observing the operation of the mechanisms used around the entrance/exit gates, known as a “sally port”. A sally port operates similarly to an airlock, with only one door ever unlocked at a time, and a guard watching via a camera to trigger the sequential locking and unlocking procedure. The purpose of this system is to prevent anybody from sneaking through, or blocking the door

open to let others escape, which works by never having a direct access path open between the inside and outside at any time (see Figure 57).

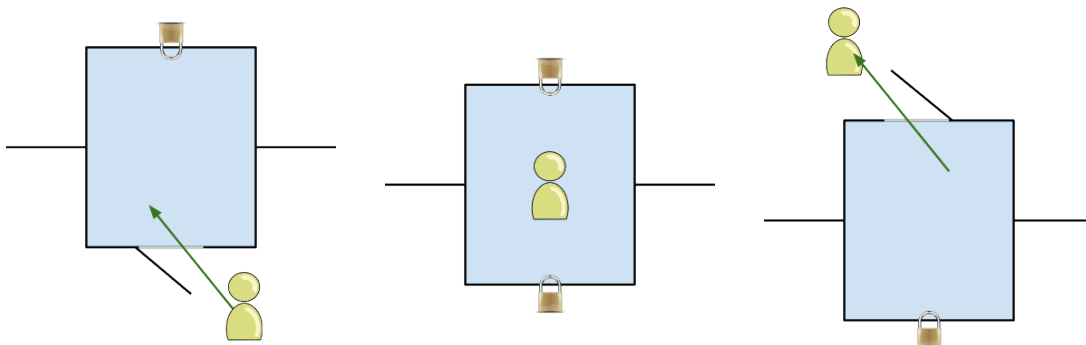


Figure 57: The sequential process of entering a secure area by means of a sally port. 1) Outer door is opened, and visitor passes into the sally port. 2) Outer door is locked. 3) Inner door is opened, and visitor exits from the sally port holding area into the secure area

One of the prisons we toured had two classrooms containing computer labs. These computers were networked through to a server in a teacher's protected office, which was in a different building, and had the KA Lite server software running on it. She also had a laptop in her office, which had Internet access, for her administrative work. The server, for security reasons (as it was networked through to the student computers) was not on the Internet, however. We realized that a "sally port" style configuration could actually work very well in this context, with the server machine being unplugged from the classroom network during downtime (e.g. on weekends), and connected instead to the Internet cable so that it could sync and download updates.

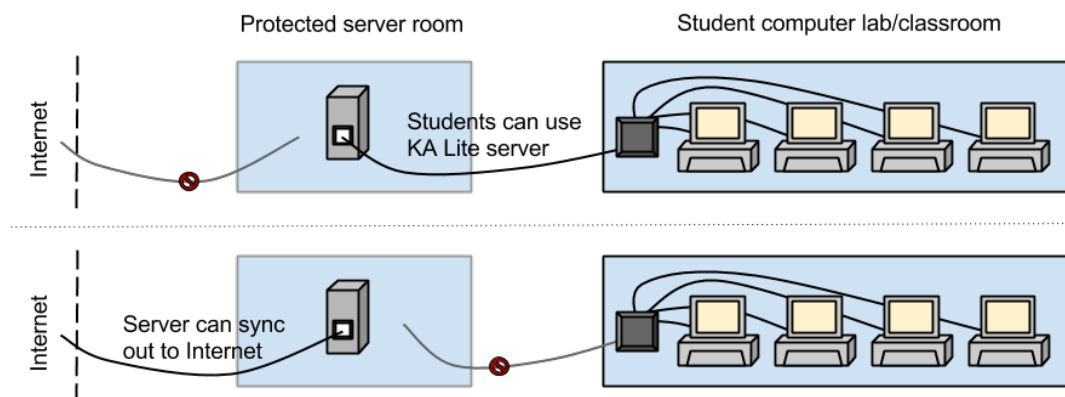


Figure 58: Digital sally port, where server is alternately connected to the classroom (upper picture), and to the Internet (lower picture)

Empowering local community-driven content creation

Education and culture

Education is one of the primary channels through which culture and communal wisdom are passed from generation to generation; children spend a significant portion of their formative years in schools, engaging with the knowledge and ideas that their societies have decided is important for them. Education is not a passive catalog of facts and skills, but an active cultural transmission process: a contract between teacher and student, elders and children, that shapes an individual's understanding of, and relationship with, the community to which she or he belongs. Viewed within this context, it becomes clear that we must approach with great care any practice involving the importation of content from curricular repositories originating outside of a community. In addition to potential issues with the content itself -- such as

language barriers, subject irrelevance, or social value mismatches -- the act of importing content and thereby deferring to an external authority runs the risk of sending a message to students of devaluing their own cultural inheritance.

At the same time, we do not want to create a siloed educational landscape, and must aim to facilitate collaboration, sharing of ideas, and deduplication of effort, so it is critical that we find ways to foster the production and distribution of educational content in a way that is inclusive and participatory. In order to accomplish this, an educational platform should put the means of producing and organizing content into the hands of the end users -- teachers, students, and other stakeholders in the educational process. Providing tools to enable users to author, remix, curate, and flexibly utilize open educational resources, and share the products of these efforts back with the broader community, has the potential to bring greater diversity, dialogue, and fairness to the global open education movement.

Above and beyond the many technical, economic, and social barriers to broad developing-world adoption of Open Educational Resources (OERs), which we have already examined, lies a more fundamental problem: the vast majority of extant OER content is written or spoken in English or other European languages, and aligned closely with Western (and particularly American) curricular standards. In the following sections, I will examine this issue of linguistic and curricular imbalance in the Open Education space, through a survey of the resources currently being made available through several of the primary online OER repositories.

Linguistic diversity of Open Educational Resources

The table in Appendix C lists the counts of content in a variety of languages on three of the main OER aggregators (OER Commons²¹, Internet Archive²², and the Open Courseware Consortium, OCW-C²³) and two of the main producers of OER content (Khan Academy²⁴ and MIT OCW²⁵), along with a measure of their linguistic diversity. The linguistic diversity index is computed using a standard metric of population diversity, “Shannon’s diversity”, or “H” (Teachman, 1980), a special case of Shannon entropy that is frequently used to measure the diversity of species with an ecosystem. The diversity scores shown in Appendix C are calculated as the entropy of the proportional distribution of materials across languages, in each of the repositories. For comparison, we estimate the global linguistic diversity score, in terms of proportions of native speakers of each language²⁶, to be approximately 5. Of all the repositories and content providers, Khan Academy comes the closest to reaching this level of diversity, with a strong emphasis on dubbing and translation into as many languages as possible, but the bulk of the translations are still primarily into European languages.

²¹ <http://www.oercommons.org/>

²² <https://archive.org/>

²³ <http://www.ocwconsortium.org/>

²⁴ <https://www.khanacademy.org/>

²⁵ <http://ocw.mit.edu/>

²⁶ Based on language counts from Nationalencyklopedin "Världens 100 största språk 2007" (The World's 100 Largest Languages in 2007), relisted at http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

Looking forwards

Khan Academy content is the first step in our push to create a distributed global network for sharing content with disconnected and low bandwidth communities. As the KA Lite platform develops, and capacities for local content curation, and incorporation of internationalized versions of content are brought into the platform, the next step will be to incorporate the capability for users to acquire and interact with other currently existing Open Educational Resources, such as Wikipedia, Connexions, and materials hosted by the OER Commons. The platform will also provide tools for the remixing and repurposing of this Open Educational content for the local, specific needs of the offline and low bandwidth communities, which can then be shared back to the broader world, and back to Open Educational Resource providers on the global Internet.

In addition, the system will enable local teachers to author their own content that can then be shared back with their local community, through the Peer to Peer Syncing mechanisms, and eventually with the global community at large. By empowering local users to create their own content, we would be enabling a latent pool of global talent to contribute to the Open Educational Resource movement, and in particular to help fill in the linguistic, cultural, and curricular content gaps that currently exist in Western-dominated OER repositories. Additionally, it affords the disconnected and low bandwidth communities using the platform the opportunity to be not just consumers, but empowered creators of content, learning skills that will help them overcome the

digital divide, and enhance educational opportunities for their communities and the world, while participating as equal partners in a global exchange of knowledge.

By providing educators of all kinds with tools to create content anywhere from plain text, images, interactive exercises, and even videos, educators will become empowered by technology in their classrooms, not replaced and marginalized. It will allow them to leverage the technology for mastery based learning for students within more enriched, relevant curricula, devised by local people to serve the needs of the local community and the students they are seeking to educate.

Chapter 5: Whither now?

While much has been accomplished in pursuit of the aims and initiatives described in this dissertation, it remains the case that many more doors have been opened than have been closed. Hence, it is worthwhile mentioning at this point some of the possible paths forwards, for our own clarity as we proceed and prioritize our next steps, and also as dangling teasers for readers who may wish to pick up on unfinished threads and weave them further into their own work. Everything described in this dissertation -- both data (pending anonymization) and code -- can be made freely available to others upon request, and will hopefully be published in a public forum for broader use in the near future.

ESL Genie

The ESL Genie project described in Chapter 2 lives on as a website at <http://eslgenie.com>, and has already produced a huge dataset that is ripe for further analysis, including data from 435 subjects with:

- Detailed language background, including first language and all other languages for which the subject has had with significant exposure; self-reported speaking, listening, writing, and reading proficiency, length and type of exposure, and proportion of each language spoken over the past year.
- For each question answered (between 100 and 347 answers per subject): Timestamp, which answer was given, correctness of answer, duration hovered over each answer, time spent regressing to previous slide, total time spent on current slide, time from last regress to click, time before final

answer was hovered prior to click, and detailed timestamped mouse movement log within the quizzing frame.

- Typed essay of 200 words or more, in response to a pictorial scenario. For the second half of the subjects, the essay was typed into Etherpad Lite, and thus includes timecourse data such as edits, deletions, pauses, etc.
- For the second half of the subjects: Self-paced reading times for 3 short pieces of text (total 791 words), two of which were stories and one of which was more academic. Some of these texts contained garden path sentences, as well as complex subject and object relative clauses, and the agreement error described in Chapter 2.

Some examples of questions that would be valuable to examine further using this dataset:

- (as originally proposed): What are the multiple independent latent traits underlying the grammatical knowledge captured by the quizzing instruments (e.g. using Multiple Item Response Theory)?
- What are the relationships between the grammatical error tags applied to each question and the latent subject traits? How do the questions cluster?
- What differences can be observed between learners on the basis of first language and years of experience? How do the subjects cluster?
- What is the relationship between the errors subjects made during free text production in the essay and the errors they (mis-)identified during the multiple choice quizzing? (This will require further manual error tagging on the essays).

- Can we identify the sequential effects, such as learning or fatigue, in the dataset? Can these be related to the time learners spent examining questions after answering them, as this might be a primary source of learning?
- Are there any more memory/CPU efficient heuristics for determining optimal question selection than calculating full posterior ability estimates?
- How can we leverage our estimates of learner ability to optimally perform learning interventions, such as providing focused overall feedback or directing the user to appropriate learning resources?

KA Lite

The KA Lite project described in Chapter 4 is under very active development and deployment by the Foundation for Learning Equality (<http://learningequality.org>), and is very open to collaborators and contributors. The open-source code lives at <https://github.com/learningequality/ka-lite>, and the team can be reached at info@learningequality.org.

As we continue to develop KA Lite and support deployment of the platform in a wide variety of environments, there are several research questions we will be examining (e.g. through Randomized Controlled Trials in India and Nigeria) to help inform our continued efforts, including:

- Does KA Lite in a classroom environment actually enhance measurable learning outcomes in comparison with classrooms not using KA Lite?

- What components of KA Lite provide the greatest learning gains, e.g. immediate feedback, video watching, motivational mechanisms like points, etc?

Some of the primary design challenges we still face with KA Lite (and successor platforms) include:

- Facilitating the creation of local content, in an offline system, to be synchronized back with the rest of the network, towards the goal of empowering local teachers and increasing the diversity of Open Educational Resources.
- Enabling peer-to-peer synchronization of both content and usage/account data between two offline installations of KA Lite, to allow for easier sneakernet distribution.
- Adaptive testing and placement tests, to help direct learners to appropriate resources, using techniques like those developed with ESL Genie.
- Adapting KA Lite to run standalone on Android devices, in the absence of a server, to allow tablets to be taken home and brought back to a server to sync.
- Development of lower-bandwidth encodings of blackboard-style video lectures, to allow for cheaper and faster distribution through constrained channels.

- Partnering with an organization developing high-quality early literacy content, to be distributed to help bootstrap pre-literate learners up to a point where they will be able to leverage other learning materials.

Our journey has only just begun, and until everyone around the world has equal opportunities to engage with a quality, basic education, we will still have work to do. We welcome anyone who wishes to join us in pursuing this mission to reach out and find ways to get involved, as it will take a large collaborative effort to help make the dream of equitable learning a reality.

Appendices

Appendix A: Tagging scheme for annotating errors in essays

Single-word

det: incorrect/missing/extra determiner/article (an, a, the, these, some...)

word: incorrect/made-up word -- include correct word like this: *remembrance*
{word:memory}

sp: spelling -- again, can include correct spelling: *wisdem* {sp:wisdom}

poss: possessive, e.g. missing 's

cap: capitalization error, whether it should be lower case or capitalized *I Am* {cap:am}

punct: punctuation error, *As explained by Daniel Gilbert* {punct:comma} *humans tend...*

prep: incorrect preposition, *lost their belongings from* {prep:to} *the war*

num: noun should be plural but is singular or vice versa

verb: verb-related error that is not agreement or tense (e.g. wrong word form for participle)

pronoun: incorrect/missing pronoun (e.g. "he acts that way because {pronoun: he} can only...")

pos: part of speech, *influenced emotional* {pos:emotionally} *and physically*

tense: incorrect tense, such as no -ed, *...he run to the car* {tense: ran}, *finally she turn* {tense:turned} *it in*

agr: agreement (e.g. subject-verb agreement: "they is")

conn: incorrect/missing connective (e.g. and, but, either, or, therefore, so, since)

rel: relativizer

Phrase

join: problem with joining words (or failing to) (e.g. hyphens or missing space between 2 words)

parallel: nonparallel structure used (e.g. how people think, behave, their memory {parallel})

omit: omit the previous word or words, *The mind is something* {omit:something} *complicated.*

order: wrong word order (e.g. "we suffer for a long time of period {order: period of time}")

frag: sentence fragment (phrase not properly attached to sentence with a connective, e.g. conjunction)

replace: {replace:new stuff}old stuff{/replace}

subcat: when verb arguments are wrong (types of prep. phrases, number/types of objects, etc)

aspect: incorrect aspect (e.g. "what people are interested in read {aspect: reading}")

do: errors involving incorrect usage of "do": e.g. "Did you thought?" or "I not think so."

comparative: "the most hugest house", or "which of the two was biggest?"

neg: negation error, for example: "made the choice to unappreciate the brilliance" -> "made the choice not to appreciate the brilliance"

Sentence

run-on: run on sentence, put at end of sentence...?

awk: awkward wording, doesn't really make sense (doesn't sound like fluent English)

gibberish: sentence is unintelligible and cannot be deciphered. Prolonged exposure likely to cause headaches.

Emptiness

insert (formerly 'missing'): word missing (e.g. "things that {verb, missing:are} happening now")

Appendix B: 84 known countries where KA Lite has been installed

Argentina	Finland	Madagascar	Russian
Australia	France	Malawi	Federation
Austria	Germany	Malaysia	Saudi Arabia
Bangladesh	Ghana	Mexico	Senegal
Belgium	Grenada	Mongolia	Serbia
Bhutan	Guatemala	Myanmar	Singapore
Brazil	Guyana	Namibia	Slovenia
Bulgaria	Haiti	Nepal	South Africa
Cameroon	Hong Kong	Netherlands	Spain
Canada	Hungary	New Zealand	Sri Lanka
Central African Republic	India	Nigeria	Sweden
China	Indonesia	Norway	Switzerland
Cook Islands	Ireland	Pakistan	Taiwan
Czech Republic	Italy	Paraguay	Tanzania
Denmark	Jamaica	Peru	Thailand
Dominican Republic	Japan	Philippines	USA (46+ states)
Ecuador	Kenya	Poland	Uganda
Egypt	Korea	Portugal	Ukraine
Ethiopia	Kuwait	Puerto Rico	United Arab Emirates
Europe	Latvia	Qatar	United Kingdom
Fiji	Liberia	Romania	Zambia
			Zimbabwe

Appendix C: Linguistic diversity and language proportions in Open Educational Resource repositories

	OER Commons Media		Internet Archive Results		Khan Academy Videos		MIT OCW Courses		OCW-C Courses	
Linguistic diversity index	1.14		1.53		3.95		1.07		1.40	
English	46098	87%	1345832	77.3%	5242	26%	3765	80.9%	18412	75.8%
Spanish	946	1.8%	25340	1.5%	2540	12%	108	2.32%	1883	7.8%
Chinese	141	0.3%	868	0.1%	740	3.6%	523	11.2%	1152	4.7%
French	693	1.3%	111736	6.4%	447	2.2%			66	0.3%
German	295	0.6%	112425	6.5%	294	1.4%			14	0.1%
Turkish	113	0.2%	1146	0.1%	1300	6.4%	38	0.82%	31	0.1%
Arabic	811	1.5%	28935	1.7%	852	4.2%				
Portuguese	133	0.3%	4320	0.3%	925	4.5%	103	2.21%	43	0.2%
Russian	100	0.2%	9316	0.5%	1170	5.7%			36	0.2%
Urdu	4	0.0%	1540	0.1%	983	4.8%				
Bengali			58	0.0%	960	4.7%				
Swahili	3	0.0%	632	0.0%	851	4.2%				
Xhosa			4	0.0%	799	3.9%				
Polish	88	0.2%	2910	0.2%	691	3.4%			87	0.4%
Italian	246	0.5%	25796	1.5%	244	1.2%			6	0.0%
Hebrew	5	0.0%	330	0.0%	609	3.0%			33	0.1%
Dutch	1217	2.3%	11896	0.7%					39	0.2%
Danish	115	0.2%	3515	0.2%	317	1.6%				
Persian	106	0.2%	260	0.0%	90	0.4%	59	1.27%	8	0.0%
Thai	22	0.0%	1741	0.1%	100	0.5%	43	0.92%		
Greek	111	0.2%	222	0.0%	253	1.2%				
Indonesian	1	0.0%	4969	0.3%	112	0.6%			8	0.0%
Korean	110	0.2%	78	0.0%	59	0.3%	13	0.28%	30	0.1%
Azerbaijani			13464	0.8%						
Serbo-Croatian	191	0.4%	1312	0.1%	69	0.3%			45	0.2%
Armenian			508	0.0%	148	0.7%				
Telugu			2492	0.1%	121	0.6%				
Czech	31	0.1%	2241	0.1%	103	0.5%				
Japanese	47	0.1%	669	0.0%	112	0.6%			332	1.4%
Bulgarian	1	0.0%	318	0.0%	129	0.6%				
Ukrainian	67	0.1%	1529	0.1%	75	0.4%			1	0.0%
Estonian	267	0.5%	142	0.0%						
Romanian	25	0.1%	5658	0.3%						
Catalan	132	0.3%	1635	0.1%					207	0.9%

Swedish	52	0.1%	4087	0.2%						
Hungarian	113	0.2%	1993	0.1%						
Norwegian	11	0.0%	907	0.1%	46	0.2%				
Vietnamese	120	0.2%	295	0.0%						
Macedonian	120	0.2%	26	0.0%						
Malay	16	0.0%	3107	0.2%					2	0.0%
Slovak	92	0.2%	418	0.0%						
Irish	97	0.2%	123	0.0%						
Finnish	53	0.1%	577	0.0%					1	0.0%
Mongolian	29	0.1%	14	0.0%	15	0.1%				
Georgian	66	0.1%	17	0.0%						
Slovenian	44	0.1%	472	0.0%					1847	7.6%
Cambodian	56	0.1%								
Basque	37	0.1%	215	0.0%						
Tamil	8	0.0%	215	0.0%	8	0.0%				
Latvian	35	0.1%								
Albanian	31	0.1%	129	0.0%						

References

- About The Licenses - Creative Commons. (n.d.). Retrieved February 24, 2014, from <http://creativecommons.org/licenses/>
- Alvarez, E. (2013, January 8). OLPC XO-4 to sell starting at \$206, production commencing March. Retrieved December 17, 2013, from <http://www.engadget.com/2013/01/08/olpc-xo-4-pricing/>
- Asimov, I. (1964, August 16). Visit to the World's Fair of 2014. New York Times, p. SM20.
- Baldwin, J. (1985). *The Price of the Ticket: Collected Nonfiction, 1948-1985*. St. Martin's Press.
- Baran, P. (1964). On Distributed Communications Networks. *Communications Systems, IEEE Transactions on*, 12(1), 1–9.
- Bray, M. (2006). Private supplementary tutoring: comparative perspectives on patterns and implications. *Compare: A Journal of Comparative and International Education*, 36(4), 515–530.
- Brown, R. (1973). *A first language: The early stages*. Harvard U. Press. Retrieved from <http://doi.apa.org/psycinfo/1973-30971-000>
- Byers, P. K. (Ed.). (2000). *Encyclopedia of World Biography: Chippendale - Dickinson* / [senior ed.: Gale, Cengage Learning. Retrieved from <http://books.google.com/books?id=J9aLmgEACAAJ>
- Cheng, S. S., Liu, E. Z. F., Ko, H. W., & Lin, C. H. (2007). Learning with online tutoring: Rural area students' perception of satisfaction with synchronous learning. *International Journal of Computers and Communications*, 1(2). Retrieved from http://www.gec.ntut.edu.tw/ezfiles/17/1017/attach/46/pta_1292_4088514_07607.pdf
- Chicago Tribune Company. (2013, December 17). \$38 tablet computer coming to U.S. Retrieved December 17, 2013, from

<http://www.chicagotribune.com/business/technology/chi-38-tablet-computer-20131217,0,6585895.story>

Compaine, B. (2001). Re-examining the digital divide. *Communications Policy in Transition: The Internet*. Retrieved from <http://books.google.com/books?hl=en&lr=&id=Lw5jBBkMeTcC&oi=fnd&pg=PA321&dq=examining+digital+divide+Compaine&ots=5wl-nXWw6z&sig=6NI64FNAXZPoCMtDpQu12r-KPyo>

Cooper, D., & Latham, M. (1967, September 20). Tomorrow's World: Home Computer Terminal 20 September 1967 - BBC [Television broadcast]. BBC Television. Retrieved from <https://www.youtube.com/watch?v=aJ6SbvrjxZA>

Davis, L., Bozick, R., Steele, J., Saunders, J., & Miles, J. (2013). *Evaluating the Effectiveness of Correctional Education: A Meta-Analysis of Programs That Provide Education to Incarcerated Adults*. RAND Corporation. Retrieved from <http://books.google.com/books?id=qQouAQAAQBAJ>

DiMaggio, P., & Hargittai, E. (2001). From the "digital divide" to "digital inequality": Studying Internet use as penetration increases. *University Center for Arts and Cultural* Retrieved from http://www.maximise-ict.co.uk/WP15_DiMaggioHargittai.pdf

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, 23(2), 245–258.

Dulay, H. C., & Burt, M. K. (1974). Natural Sequences in Child Second Language Acquisition. *Language Learning*, 24(1), 37–53.

Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92(4), 605.

Ellis, N., & Laporte, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. *Tutorials in Bilingualism: Psycholinguistic Perspectives*.

Global tutoring industry experiencing explosive growth - ICEF Monitor - Market intelligence for international student recruitment. (n.d.). Retrieved February

20, 2014, from
<http://monitor.icef.com/2012/11/global-tutoring-industry-experiencing-explosive-growth/>

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “Natural Order of L2 Morpheme Acquisition” in English: A Meta-analysis of Multiple Determinants. *Language Learning*, 51(1), 1–50.

Groves, M., Régo, P., & O'Rourke, P. (2005). Tutoring in problem-based learning medical curricula: the influence of tutor background and style on effectiveness. *BMC Medical Education*, 5(1), 20.

Gupta, A. (Ed.). (2007). *Going to School in South Asia*. Santa Barbara, CA: Greenwood. Retrieved from <http://ebooks.abc-clio.com/reader.aspx?isbn=9780313088773>

Hagen, L. (1994). Unification-based parsing applications for intelligent foreign language tutoring systems. *Calico Journal*. Retrieved from <http://journals.sfu.ca/CALICO/index.php/calico/article/view/482>

Harper, T. (2013, November 13). Android Tops 80% Global Smartphone Market Share – Windows Phone up 156% Year on Year. Retrieved December 17, 2013, from <http://www.highlightpress.com/android-tops-80-global-smartphone-market-share-windows-phone-up-156-year-on-year/6708/tharper>

Hartley, S. (1977). Meta-analysis of the effects of individually paced instruction in mathematics.

Heift, T., & Schulze, M. (2007). Errors and intelligence in computer-assisted language learning: parsers and .. (p. 283). Routledge. Retrieved from <http://books.google.com/books?id=OUZ8s8B1pIUC&pgis=1>

Hilbert, M. (2010a). When is Cheap, Cheap Enough to Bridge the Digital Divide? Modeling Income Related Structural Challenges of Technology Diffusion in Latin America. *World Development*, 38(5), 756–770.

Hilbert, M. (2010b). When is Cheap, Cheap Enough to Bridge the Digital Divide? Modeling Income Related Structural Challenges of Technology Diffusion in Latin America. *World Development*, 38(5), 756–770.

History - Creative Commons. (n.d.). Retrieved February 24, 2014, from <http://creativecommons.org/about/history/>

Izumi, S. (2003). Processing Difficulty in Comprehension and Production of Relative Clauses by Learners of English as a Second Language. *Language Learning*, 53(2), 285–323.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25(04), 603–634.

Kamenetz, A. (2013, October 2). Idaho Prisoners Use Khan Academy Offline to Study for GEDs. Retrieved from <http://www.fastcolabs.com/3018668/idaho-prisoners-use-khan-academy-offline-to-get-geds>

Khan, S. (2012). *The One World Schoolhouse: Education Reimagined*. Grand Central Publishing. Retrieved from <http://books.google.com/books?id=xz-gkDYm4UUC>

Krashen, S. (1982). *Principles and practice in second language acquisition*. aces.ir. Retrieved from <http://aces.ir/attachments/22d1286622494-communicative-approach-stephen-crashen.pdf>

Kravets, D. (2013, October 14). NSA Leaks Prompt Rethinking of U.S. Control Over the Internet's Infrastructure. Retrieved February 24, 2014, from <http://www.wired.com/threatlevel/2013/10/global-net-infrastructure/>

Lage, M., Platt, G., & Treglia, M. (2000). Inverting the Classroom: A gateway to Creating an Inclusive Learning Environment. *Journal of Economic Education*. Retrieved from http://dl.dropboxusercontent.com/u/249331/Inverted_Classroom_Paper.pdf

- Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2013). *Ethnologue: Languages of the World*, Seventeenth edition. Dallas, Texas: SIL International. Retrieved from <http://www.ethnologue.com>
- Lightbown, P. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In *Second Language Acquisition*.
- Looms, P. O. et al. (2011). *Making Television Accessible*. Retrieved from http://www.itu.int/ITU-D/sis/PwDs/Documents/ITU-G3ict%20Making_TV_Accessible_Report_November_2011.pdf
- Marill, T., & Roberts, L. G. (1966). Toward a Cooperative Network of Time-shared Computers. In *Proceedings of the November 7-10, 1966, Fall Joint Computer Conference* (pp. 425–431). New York, NY, USA: ACM.
doi:10.1145/1464291.1464336
- Miller, M. (2013, June 4). Smartphone avg selling price drops while growth reaches 33%. Retrieved December 17, 2013, from <http://www.zdnet.com/smartphone-avg-selling-price-drops-while-growth-reaches-33-7000016347/>
- MITnews. (2005, November 16). Annan presents prototype \$100 laptop at World Summit on Information Society. Retrieved from <http://web.mit.edu/newsoffice/2005/laptop-1116.html>
- Open Handset Alliance. (2007, November 5). Industry Leaders Announce Open Platform for Mobile Devices. Retrieved December 17, 2013, from http://www.openhandsetalliance.com/press_110507.html
- Rose, P., Akyeampong, K., Antoninis, M., Barry, M., Bella, N., Chemery, E., ... Zubairi, A. (2014). *Education for All Global Monitoring Report 2013/14; Teaching and Learning: Achieving quality for all*. Retrieved from <http://unesdoc.unesco.org/images/0022/002256/225660e.pdf>
- Smith, M. S. (2013). Open Educational Resources: Opportunities and Challenges for the Developing World. In Smith, Matthew L. and Reilly, Katherine M. A. (Ed.), *Open Development: Networked Innovations in International Development* (pp. 129–170). MIT Press.

- Tansel, A., & Bircan, F. (2006). Demand for education in Turkey: A tobit analysis of private tutoring expenditures. *Economics of Education Review*, 25(3), 303–313.
- Teachman, J. D. (1980). Analysis of Population Diversity: Measures of Qualitative Variation. *Sociological Methods & Research*, 8(3), 341–362.
- The Great Didactic of John Amos Comenius. (1907). London: Adam and Charles Black. Retrieved from https://archive.org/stream/cu31924031053709/cu31924031053709_djvu.txt
- Uche, L. U. (1989). *Mass Media, People, and Politics in Nigeria*. Concept Publishing Company. Retrieved from <http://books.google.com/books?id=bX7aJ8faPvoC>
- UN General Assembly. (1948). Universal declaration of human rights, Article 26. 217 A (III). Retrieved from <http://www.un.org/en/documents/udhr/index.shtml#a26>
- UN General Assembly. (1966). International Covenant on Economic, Social and Cultural Rights. 2200A (XXI). Retrieved from <http://www.ohchr.org/EN/ProfessionalInterest/Pages/CESCR.aspx>
- United Nations. (2013). United Nations Millennium Development Goals: Achieve Universal Primary Education. Retrieved February 23, 2014, from <http://www.un.org/millenniumgoals/education.shtml>
- Upton, L. (2013, November 17). TWO MILLION! | Raspberry Pi. Retrieved February 24, 2014, from <http://www.raspberrypi.org/archives/5265>
- Van Zuilen, A. (1977). *The Life Cycle of Magazines: A Historical Study of the Decline and Fall of the General Interest Audience Magazine in the United States During the Period Uithoorn [the Netherlands]: Graduate Press.*
- VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197–221.

VanPatten, B., Keating, G. D., & Leiser, M. J. (2012). Missing verbal inflections as a representational problem: Evidence from self-paced reading. *Linguistic Approaches to Bilingualism*, 2(2), 109–140.

Vest, C. M. (2001). Disturbing the Educational Universe: Universities in the Digital Age -- Dinosaurs or Prometheans? Retrieved February 24, 2014, from <http://web.mit.edu/president/communications/rpt00-01.html>

Warschauer, M. (1996). Computer-assisted language learning: An introduction. *Multimedia Language Teaching*.

Zuckerberg, M. (2013). Is Connectivity A Human Right? Retrieved from https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-ash3/851575_228794233937224_51579300_n.pdf