The State of Wiki Usage in U.S. K-12 Schools: Leveraging Web 2.0 Data Warehouses to Study Quality and Equality in Online Learning Environments

Blair Justin Fire Reich

Richard Murnane
John Willett
Hunter Gehlbach

A Thesis Presented to the Faculty
of the Graduate School of Education of Harvard University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

2012

©2012
Blair Justin Fire Reich
Some Rights Reserved
This work is licensed under a Creative Commons Attribution 3.0 Unported License
(CC BY 3.0)
http://creativecommons.org/licenses/by/3.0/

For Sheryl, Elsa,and Adella

Acknowledgements

This research has been built on the work and generosity of hundreds of people:

- The Hewlett Foundation's Open Education Resources initiative, which generously provided the bulk of the funding for this research.
- The hundreds of educators and students who participated in this research by sharing their work online, by taking time to speak with us, or by inviting us into their classroom.
- The small army of research assistants who worked on this project: Sara Garcia, Sarah Carruci, Shane Tutwiler, Noriko Magari, Heather Zhuang, Edwin Acosta, Louis Choi, Seth Woodworth, Amy Hoffmaster, Eric Gustavson, Jennifer Worden, Natasha Joshi, Stephanie Chronenberg, Sara Park, Moise Derosier, Lindsay Poland, Jessica Houser, Alexis Morgan, James Cody, and Sabita Verma. Special thanks to Bev Boos for her long service.
- PBworks.com, who provided data for this research, though no financial compensation.
- Bob Wolf, who helped develop the foundation of the research project and provided valuable guidance in the early research design
- My colleagues in three courses—Applied Longitudinal Data Analysis, Discourse Analysis, and Logics of Research Design—who read various proposals and papers related to this work. Special thanks to the professors in those courses, John Willett, Helen Haste, and Jal Mehta for their feedback.
- Stone Wiske, who supervised an independent study and provided feedback on various parts of this project
- Meira Levinson has read sections of this paper, listened to my talks, read preliminary proposals, and has been my confidant and friend throughout my doctoral career.
- Steven Yussen and the three anonymous reviewers from *Educational Researcher* who provided feedback on the first section.
- My committee of readers, Hunter Gehlbach, Richard Murnane, and John Willett provided nuanced feedback, kind words of encouragement, careful proofreading, and plenty of spirited and humorous conversation. Very special thanks to Dick, for trusting me and supporting me in writing a grant and building a lab.
- To three generations of women in my life: to Sheryl, who taught me to believe in the capacity of young people; to Elsa, for being my model of scientific integrity as well as my best friend; and to Adella, whose very existence motivated the completion of this document.

Dissertation Abstract

In the first part of this dissertation, I document wiki usage in U.S. K-12 settings by analyzing data on a representative sample drawn from a population of nearly 180,000 wikis. My research group, which I lead and managed, measured the opportunities wikis provide for students to develop 21st century skills such as expert thinking, complex communication, and new media literacy. There are four types of wiki usage: (1) trial wikis and teacher resource-sharing sites (40%), (2) teacher content-delivery sites (34%), (3) individual student assignments and portfolios (25%) and (4) collaborative student presentations and workspaces (1%). Wikis created in schools serving low-income students have fewer opportunities for 21st century skill development and shorter lifetimes than wikis from schools serving affluent students. In this study, I illustrate the exciting potential that Web 2.0 data warehouses offer for educational research. In an extensive methodological addendum, I describe how the data from wikis were gathered and how wiki quality was assessed. The first part of this dissertation was published in *Educational Researcher*.

In the second part of this dissertation, I detail the development of the *Wiki Quality Instrument* (WQI) and its related protocols. I first present the WQI, and then I describe the protocols and training procedures that our group used in applying the WQI. I then describe the development of the instrument and its associated protocols. Finally, I suggest ways that the WQI can be adapted for use by educators and other researchers in a variety of settings.

A Note on the Use of “We”

I have chosen to write this dissertation using the first person plural, “we.” In the course of this project, I managed a large team of researchers including our senior advisors and 25 research assistants. While I was the primary architect of this research, it was built as a team. It seemed appropriate, therefore, to refer to “we” throughout the dissertation.

The other practical rationale for this decision was that all of the pieces of this dissertation will be published in settings where I am the first author, and Richard Murnane and John Willett are additional authors. It did not make sense therefore, to write the dissertation using the first person singular and then re-write everything in the first person plural.

As the social sciences become more complex and interdisciplinary, I hope that the stigma attached to collaboratively authored works continues to wane and that doctoral students can complete their apprenticeship through collaborative research projects.

Table of Contents

Contents

The State of Wiki Usage in U.S., K-12 Schools: Leveraging Web 2.0 Data Warehouses to Assess Quality and Equity in Online Learning Environments 4

Abstract 4

Introduction 5

Background and Context 7

21st Century Skills and Measuring Wiki Quality 8

Web 2.0 and the Digital Divide 9

Research Questions 10

Research Design 11

Patterns of Wiki Usage 15

To What Extent Do Wikis Created in U.S. Public Schools Provide Opportunities for 21st Century Skill Development? 18

Do Wikis Created in Schools that Serve More Affluent Populations Provide More Opportunities for 21st Century Skill Development than Wikis Created in Schools Serving Less Affluent Populations? 24

Quality and Equity in U.S. K-12 Wiki Learning Environments 25

A New Direction for Education Research 28

References 31

The State of Wiki Usage in U.S., K-12 Schools: 37

Supplemental Materials on Research Design 37

Dataset 37

Sample 39

Sample Limitations 43

Instrument 46

Procedures 48

Measures 55

Outcome Measure 56

Question Predictors 56

Additional Predictors 57

Data-Analytic Strategy 58

Descriptive Statistics 58

RQ#1 To what extent do wikis provide opportunities for 21st century skill development in U.S. public schools? 58

RQ#2 Do wikis created in schools that serve more affluent populations provide more opportunities for 21st century skill development than wikis created in schools serving less affluent populations? 60

Conclusion 61

References 62

The Wiki Quality Instrument: An Instrument for Measuring Opportunities to Develop 21st Century Skills in Wiki Learning Environments 63

Introduction 65

Part I: The Wiki Quality Instrument 68

Section 1: Summary Questions 68

Section 2: Definitions 70

Section 3: Decision Rules 72

Demographic Questions 72

Quality Questions 81

Final Overall Rating Questions 88

Part II: Content Analysis and Training Guidelines and Protocols 89

Section 1: Overview of Content Analysis Protocols 89

Section 2: Wiki Sampling 90

Section 3: Content Analysis Work Flow 92

Section 4: Evaluating Wiki-Edit Histories 104

Section 5: Ethical Concerns in Online Content Analysis and IRB Considerations 108

Section 6: Overall Ratings 110

Section 7: Training protocols 111

Section 8: Conclusion 114

Part III: On the Development of the WQI 116

Defining the Context of our Study 117

Developing a Theory of Wiki Quality 123

Iterating towards the Final Wiki Quality Instrument 134

Summarizing the Design Process for the Wiki Quality Instrument 139
Part IV: On Developing Protocols for the WQI 141
How Often Should We Measure Wiki Quality? 142
When in a Wiki's Lifecycle should we take Quality Measurements? 149
The Protocol for Measuring Wiki Quality 153
Part V: Adaptation Guidelines for Educators 157
Background on the WQI 158
Using the WQI to assess wiki usage in schools and districts 160
Adapting the Wiki Quality Instrument for Classroom Use 165
Part VI: Adaptation Guidelines for Researchers 169
Applying the Current Version of the WQI in Alternate Settings 169
Modifying the WQI to Study other Wiki Providers 169
Modifying the WQI to study Wiki Usage in Particular Curricular Domains 170
Adapting the WQI for the Study of Other Social Media Platforms in K-12 Settings 173
Conclusion 175
References 177
VITA 180

The State of Wiki Usage in U.S., K-12 Schools: Leveraging Web 2.0 Data Warehouses to Assess Quality and Equity in Online Learning Environments

Abstract

To document wiki usage in U.S. K-12 settings, this study examines a representative sample drawn from a population of nearly 180,000 wikis. We measured the opportunities wikis provide for students to develop 21st century skills such as expert thinking, complex communication, and new media literacy. We find four types of wiki usage: (1) trial wikis and teacher resource-sharing sites (40%), (2) teacher content-delivery sites (34%), (3) individual student assignments and portfolios (25%) and (4) collaborative student presentations and workspaces (1%). Wikis created in schools serving low-income students have fewer opportunities for 21st century skill development and shorter lifetimes than wikis from schools serving affluent students. This study illustrates the exciting potential that Web 2.0 data warehouses offer for educational research.

Introduction

Web 2.0 tools—online platforms that allow non-programmers to contribute content to the World Wide Web (O'Reilly, 2005)—are transforming our society. Comment forums turn newspaper columns into conversations, marketers use blogs to get real-time feedback from consumers (Li & Bernoff, 2008), and dictators have fallen as dissidents organize online (Zuckerman, 2010). For good and for ill, Facebook and MySpace have reshaped notions of community, friendship, and identity for their users (boyd, 2008; Turkle, 2011). As over 2 billion Internet users share words, images, and videos through Web 2.0 tools, global changes have followed (Reich, 2008).

Web 2.0 tools have made inroads into the U.S. education system as well. In a 2009 Fast Response System Survey (FRSS) conducted by the National Center for Education Statistics (NCES), 38% of public school teachers reported using blogs or wikis for class preparation and instruction, and 21% reported that they required their students to contribute to blogs or wikis (Gray, Thomas, & Lewis, 2010). Given that these technologies have only been widely available for a decade, this represents a striking pattern of growth.

How then, can these new technologies be used within classrooms? Consider the Flat Classroom Project (www.flatclassroomproject.org), an international wiki project started by Vicki Davis of the Westwood Schools in Georgia and Julie Lindsay of the Beijing International School. Now in its sixth year, the Flat Classroom Project has engaged dozens of classrooms around the world in producing wiki pages devoted to explaining the ten “world-flatteners” described in Thomas Friedman’s *The World is Flat.*

Students work in international teams to collaboratively create a multimedia presentation about their topic. Project guidelines ensure that students work closely with their foreign peers; for instance, each team's wiki page includes videos produced with raw film recorded in one country that has been "outsourced" for editing. Thus, a video about social networks might be shot in Shanghai and edited in Vienna. Each final project presents a global student perspective on the future of technology and society. Through these experiences, students have the chance to exercise their skills in communication, creativity, collaboration, and critical thinking in a global context.

The Flat Classroom Project stands out as a proof of concept that Web 2.0 learning environments can facilitate rich educational experiences that prepare students to thrive in a networked world. But, is this student-centered, global, collaborative project representative of typical wiki usage or is it a relatively rare exception? Are such exemplary projects found in diverse settings, or do they exist primarily in schools serving affluent students? In this paper, we address these key questions of quality and equity by adopting a novel research approach made possible by the data records of Web 2.0 platforms.

Every time a user makes a change to a blog, wiki, or content management system (e.g. Blackboard or Moodle), the data warehouse supporting that online environment records the change. In aggregating these continuously- recorded actions, researchers can examine global patterns of online teaching and learning at gradations of nearly infinite granularity. We can conduct focused studies of a student's activity over a period of minutes, or we can compare hundreds of thousands of learning communities over years.

In this study, we leverage these new datasets to understand patterns of wiki usage in U.S., K-12 settings. We focus on wikis because they are emblematic of Web 2.0: they are collaborative, multimedia spaces where any community member can edit any page at any time (Larusson & Alterman, 2009; Reich & Daccord, 2008; Richardson, 2008). We analyzed hundreds of wikis randomly drawn from a population of nearly 180,000, publicly-viewable, education-related wikis. We measured wiki quality through examining the detailed edit histories of each of our sampled wikis, using a quality rating rubric that we created. We assessed equity by examining how quality differed across wikis created in schools serving different socioeconomic populations.

Our findings suggest that wikis do support emerging models of innovative, online pedagogies that can foster the development of essential competencies for a networked age. We also find that two stubborn challenges of education technology persist into the Web 2.0 era: 1) the difficulty of supporting teachers in using new tools for innovation rather than gaining efficiencies in established routines and 2) the disturbing trend of innovations taking root primarily in already-advantaged school settings. From a methodological perspective, we demonstrate that new forms of educational data allow detailed characterization of classroom technology and pedagogy at a national scale.

Background and Context

In their article asking "Web 2.0 and classroom research: what path should we take *now*?", Greenhow, Robelia and Hughes (2009) argued that researchers should focus on *what learners do* with Web 2.0 tools and *issues of equity in* networked learning environments. We concur that learner activity and equity are central themes for inquiry into a medium defined by its capacity for broadening participation in knowledge production. In this section, we describe two theories that have informed our inquiry into wiki activity and equity: *21st century skills* and the *digital divide*.

21st Century Skills and Measuring Wiki Quality

While education technology is often used to generate efficiencies in content delivery and testing, many educational technologists focus on developing higher-order thinking skills and allowing students to rehearse for future performances in a technology-rich workforce and civic sphere. Papert's (1980) work on computers and constructivism represents one important intellectual wellspring of this tradition. Scardamalia and Bereiter's (1993) work on knowledge-building communities, especially their wiki-like *Knowledge Forum* platform, represents another vein of theory that explains how learning technologies can prepare students for the challenges of life rather than for achievement tests. While, as we shall demonstrate, wikis can be used to facilitate content delivery, we are keenly interested in uses of wikis, like the Flat Classroom Project, that use the collaborative, multimedia affordances of wikis to allow students to create sophisticated performances of understanding (Wiske, Franz, & Breit, 2005).

Recent research into the skill demands of America's rapidly changing labor market has clarified the kinds of higher-order thinking skills that today's students will need to thrive in the future. Levy and Murnane (2004) provide the empirical foundation for the movement to reorient schools towards the teaching of *21st century skills*. They argue that computers have taken over many of the repetitive tasks that dominated the 20th century economy. Therefore, education should focus on developing skills where humans have a comparative advantage over computers in the 21st century labor market. Levy and Murnane identify *expert thinking* (ill-structured problem solving) and *complex communication* as the most important of these skill domains. Jenkins (2009) has proposed another compelling dimension of 21st century skills: *new media literacy,* defined as the capacity to critically consume and produce technology-rich media in a social context. While other scholars have provided their own lists of 21st century skills, Dede (2010) found that expert thinking, complex communication, and technology literacy are the key domains anchoring the prominent compilations.

This scholarship on 21st century skills provides a theoretical framework for considering how best to measure the quality of online learning environments where higher-order thinking skills are emphasized. Numerous studies have investigated the use of online environments to develop particular dimensions of higher-order thinking, such as cognitive engagement (Oriogun, Ravenscroft, & Cook, 2005), collaboration (Cortez, Nussbaum, Woywood, & Aravena, 2009; Trentin, 2009), or knowledge-building (Moskaliuk, Kimmerle, & Cress, 2009). Most of these studies have been conducted within a single subject domain, such as algebra (Chiu, 2008) or business ethics (Jeong, 2003). While these focused studies spotlight selected dimensions of online learning, in this study we attempt to build upon this research by assessing wiki-based 21st century learning broadly and at scale.

Web 2.0 and the Digital Divide

In recent decades, the profound impact of technology on the workforce and civic sphere has given rise to serious concerns about the *digital divide*: inequities in technology-rich educational opportunities. Most early investigations of the digital divide focused on issues of *access* to computing technology and raised questions about the number of computers or speed of networks in schools serving communities with differing income levels (Warschauer & Matuchniak, 2010). Attewell (2001), however, argued that issues of access were secondary to inequities in technology *usage*. Several studies that predate the Web 2.0 revolution discovered that students from economically-advantaged families were more likely to use technology for higher-order thinking with more adult involvement (Attewell & Battle, 1999; Attewell, 2003; Wenglinsky, 1998). In contrast, students from disadvantaged families tended to use computers for unsupervised drill-and-practice routines. Attewell (2001) characterized this division as the *second digital divide*, the gap between how learners use technologies in different communities. For instance, in the 2009 FRSS study cited previously, there were essentially no differences in teachers' adoption of blogs and wikis across schools serving different populations. As we shall see, however, when we examine wikis in terms of the opportunities they provide for students to develop 21st century skills, important divisions become apparent.

Research Questions

In this study, we define wiki quality in terms of the opportunities that wikis provide for students to develop 21st century skills such as expert thinking, complex communication, and new media literacy. We measure wiki quality through a detailed analysis of the edit histories of a representative sample of wikis created in U.S., K-12 public schools. Using these quality measurements, we address two questions: 1) To what extent do wikis created in U.S. public school provide opportunities for the development of 21st century skills? 2) Do wikis created in schools that serve more affluent populations provide more opportunities for the development of 21st century skills than wikis created in schools serving less affluent children?

Research Design

For this study, we drew samples from the population of all 179,851 publicly-viewable, education-related wikis hosted by PBworks.com from the founding of the site in June 2005 through August 2008. PBworks is one of the three most-visited sites that offer free wiki hosting (Alexa, 2010).These wikis are used from elementary through graduate schools across the world in nearly every academic subject. For each of these wikis, PBworks preserves every revision of every page. We summarize our methods here, and further details are available in the next chapter, "Supplementary Materials on Research Design."

From the population of 179,851 wikis, we drew a 1% random sample of 1,799 wikis and separated out the 255 wikis identifiably associated with a specific U.S., K-12 public school. Our sample includes wikis created in schools from 41 of the 50 states.[1]

To evaluate the degree to which wikis provide opportunities for students to develop 21st century skills, we applied a newly developed instrument called the Wiki

[1] In the online supplement, we discuss two potential limitations of our sample. If the wikis created at PBworks differ from the wikis hosted at other sites, such as Wikispaces.com, then we cannot generalize our findings beyond the population of PBworks wikis. We have no reason to believe that such differences exist, but further research is needed to verify this assertion. Similarly, the population of 179,851 wikis from which we drew our analytic samples represented only 70% of the education-related wikis that were created at PBworks during the time-period in question; the remaining 30% were "private" wikis and were not viewable publicly. If publicly-viewable wikis differ from private wikis, then our findings cannot be generalized to private wikis.

Quality Instrument (WQI). We developed the WQI over a two year period after 68 interviews with wiki-using teachers, 40 student focus groups, observations in 19 classrooms, and several rounds of pilot testing and revision (Reich, Murnane, & Willett, 2010). The WQI has 24 items in five subsections: (a) Information Consumption (2 items) (b) Student Participation (4 items), (b) Complex Communication (7 items), (c) Expert Thinking (5 items), and (d) New Media Literacy (6 items). In each section, coders assess whether students participate in activities that provide opportunities to develop 21st century skills: activities such as co-creating a shared page, reflecting on a work product, or embedding multimedia frames into a page. In Table 1, we present a summary list of questions posed by the WQI. (The WQI instrument and documentation describing its development and use are presented in the subsequent chapters of this dissertation.)

Table 1: *Summary questions of coding categories used in the Wiki Quality Instrument*

Category *Sub-category*	**Summary question**
Information Consumption	
Course Materials	Do students come to the wiki to access academic materials?
Information Gateway	Do students come to the wiki to access links to other Web sites?
Student Participation	
Contribution	Does at least one student contribute, in any form, to the wiki?
Individual Pages	Does at least one student own their own page on the wiki?
Shared Pages	Does at least one pair (or group) of students own their own wiki page?
Ownership	Do student(s) serve as primary facilitator and content creator of the wiki?
Expert Thinking	
Academic Knowledge	Does at least one student complete a task requiring academic knowledge (as opposed to simply writing about hobbies or one's family)?
Information Organization	Does at least one student complete a task requiring information organization, rather than routine information retrieval?
Metacognition	Does at least one student reflect on his/her work product or process?
Crediting	Does at least one student credit their sources of his/her work?
Teacher Feedback	Do teachers provide feedback on student work?
Complex Communication	
Concatenation	Do multiple students add discrete sections of text to the same page?
Copyediting	Does at least one student copyedit text created by another student?
Co-Construction	Does at least one student substantively edit text created by another student?
Commenting	Does at least one student comment upon another student's work on the wiki?
Discussion	Do students respond to each others' comments for at least four conversational turns?
Scheduling	Do students schedule meetings or tasks?
Planning	Do students plan for future work?
New Media Literacy	
Formatting	Does at least one student use formatting elements beyond plain text?
Links	Does at least one student post a link to another page or document?
Hyperlinks	Does at least one student create links rendered as simple text or images?
Images	Does at least one student embed an image into a page?
Uploads	Does at least one student upload a document?
Multimedia	Does at least one student embed a multimedia element into a page?

The WQI poses dichotomous questions about the presence or absence of activities that can provide students with opportunities to develop 21st century skills. We do not compare the quality or the frequency of activities between wikis, as the learning environments in our sample are too diverse for scalar comparisons (such studies would be fruitful within narrower domains, such 7th grade Earth Science wikis.). Nor do we measure actual student development, as we cannot assess baseline competencies or changes in student competencies resulting from wiki-based learning (such studies would be fruitful if students could be assessed outside the wiki). Instead, the WQI measures "opportunities for 21st century skill development," a set of behaviors which are pre-conditions for 21st century skill development.

Each of our 255 wikis was coded by two research assistants, and then reconciled by a third senior research assistant. Our coders identified several key features of each wiki: its academic subject area(s), student grade level(s), and host school. From these school names, we obtained each school's Title I eligibility and percentage of students eligible for Free and Reduced Priced Lunches (FRPL) as indicators of school-level socioeconomic status (National Center for Education Statistics, 2007-2008). To create a final outcome measure, we summed the values of our 24 dichotomous WQI items to form a 0-24 point composite wiki quality scale. Interrater agreement across our 24 items averaged .92.

We measured wiki quality at 7, 14, 30, 60, 100, and 400 days. We found that wiki quality trajectories tended to be logarithmic; typically, wiki quality rises quickly within the first two weeks and then the quality trajectories level off. As a result, we use wiki quality at day 14 as a summary statistic that permits consistent comparison across wiki learning communities.

To address our first research question concerning overall quality levels in U.S., K-12 wikis, we present the distribution of composite wiki quality scores in our sample as well as descriptions of wikis at various quality levels. To test whether quality differs in wikis created in socioeconomically different schools, we use Poisson regression, since our wiki quality scores have features in common with count data (we count the number of identifiable wiki behaviors that can promote 21st century skill development). We fit a Poisson regression model with the composite wiki quality score as the outcome and the percentage of students eligible for FRPL in a wiki's school as our question predictor.

Patterns of Wiki Usage

In our analysis of wiki-using classrooms, we found an extraordinary diversity of learning activities. Students used wikis to publish homework assignments, maintain portfolios, peer review writing, post artwork, download music for rehearsals, and review drills for physical education. One sampled wiki began as a teacher-facilitated reading group for middle-school girls, and the girls voluntarily continued using the wiki to maintain a detailed table mapping romantic crushes within their grade.

This diverse activity occurred throughout the K-12 sector. Of our 255 public school wikis, 25% supported instruction in grades K-5, 28% in grades 6-8, and 52% in grades 9-12 (the sum of these percentages exceeds 100% because some wikis supported multiple grades). Wikis were used not just in computer classes; they supported instruction throughout the curriculum. We found that 34% of wikis supported English/language arts instruction, 13% supported social studies, 18% supported science, 13% supported math, 14% supported computer science, and 26% supported another subject or no subject. These findings generally align with the distributions of wiki adoption reported in the aforementioned 2009 FRSS survey.

In addition to these cross-sectional data, we measured each wiki's lifetime from the moment of creation until the final page edit. In Figure 1, we present the Kaplan-Meier estimated survivor function for our wiki sample (Singer & Willett, 2003). We display the time since wiki creation on the X-axis and estimated survival probabilities (the proportion of wikis that remain active beyond each particular time-point) on the Y-axis.

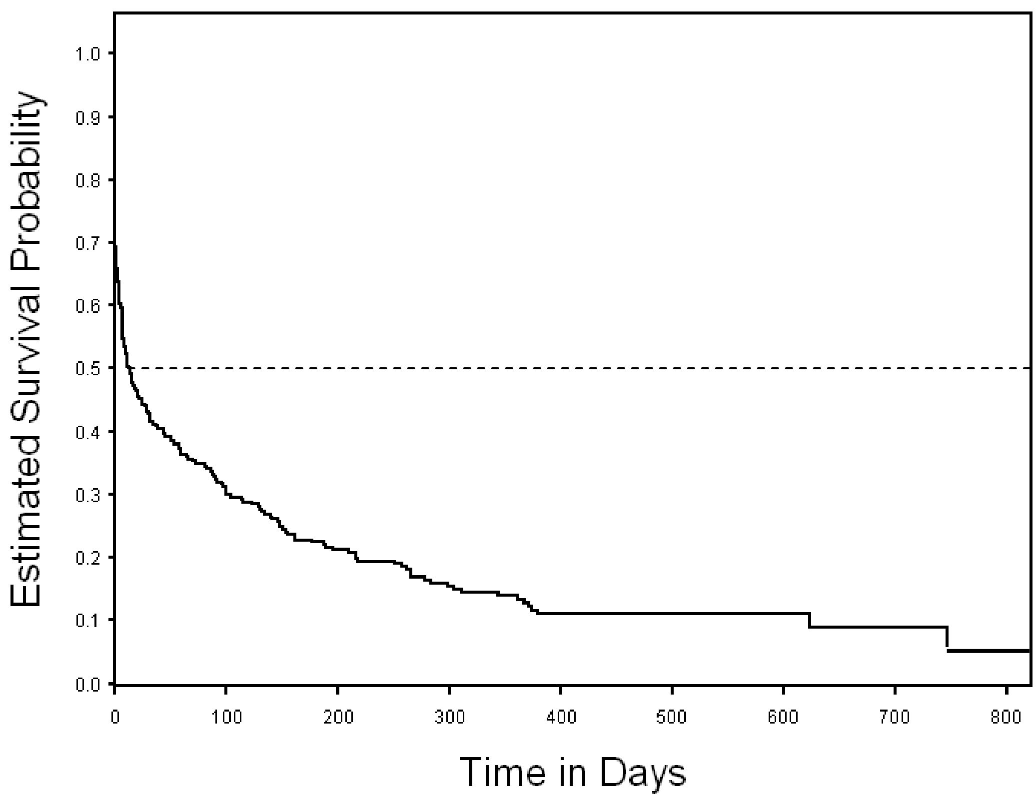

Figure 1. Estimated survivor function of wikis created in U.S. public schools (n=255).

The steep initial drop in the estimated survivor function indicates that many wikis are terminated almost immediately after creation. For instance, the estimated median lifetime (the length of time beyond which 50% of the original wikis survive) of public school wikis is only 13 days, and only one quarter of wikis persist beyond 151 days. These estimates suggest that most wikis that are used at all are used for short-term projects and assignments rather than serving as long-term course platforms or student portfolios.

We also found evidence that wikis created in schools serving predominantly low-income families cease development earlier than wikis created in other schools. In Figure 2, we display estimated survivor functions for wikis created in Title I eligible versus non-Title I eligible schools. On average, wikis created within non-Title I eligible schools persist longer (Wilcoxon χ^2=11.38, *df*=1, p=0.0007). The estimated median lifetime for wikis created in non-Title I schools is 32 days compared to 6.5 days for wikis created in Title I schools. Furthermore, 42% of wikis created within Title I schools do not last more than 1 day, compared to 21% of wikis created within non-Title I schools. Longevity is by no means a perfect proxy for wiki quality, but these findings provide one indication of the second digital divide of usage.

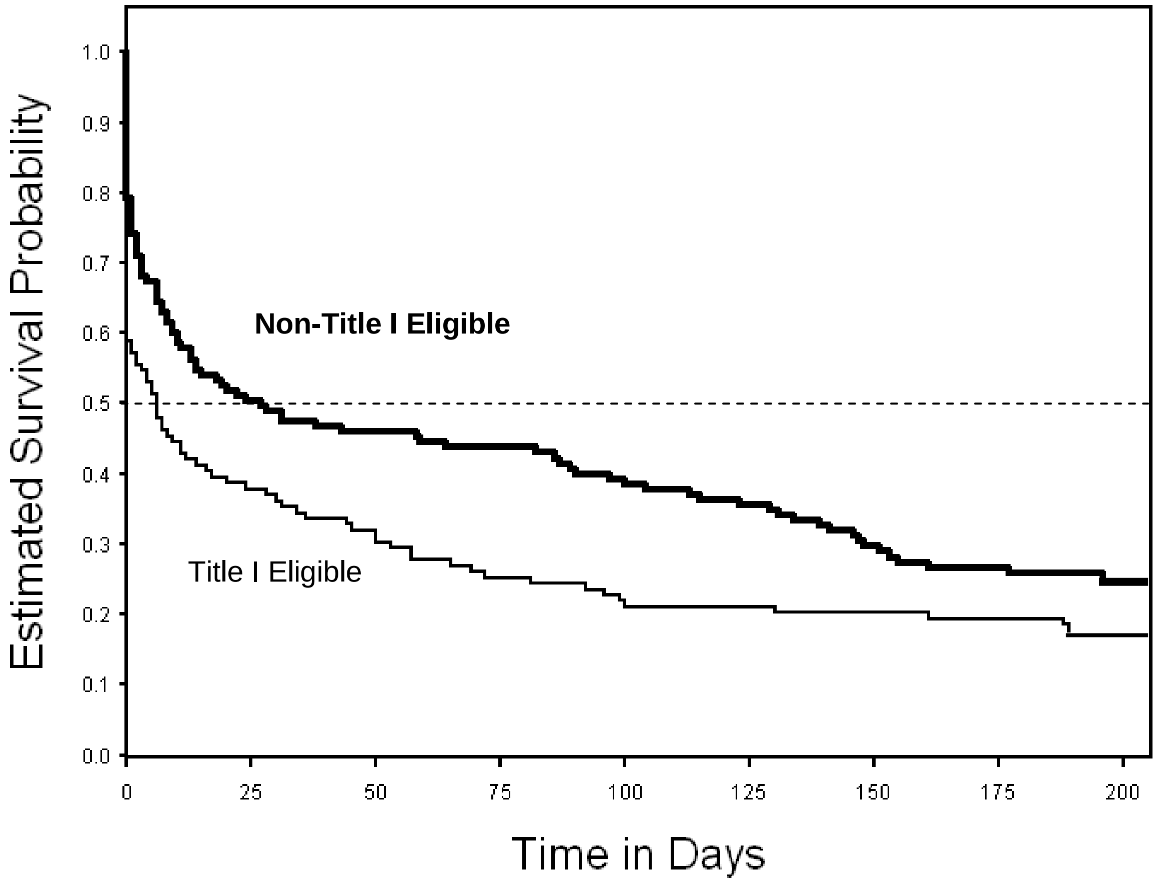

Figure 2. Estimated survivor functions for wikis hosted by Title I eligible (n=117) and non-Title I eligible schools (n=133), through day 200.

To What Extent Do Wikis Created in U.S. Public Schools Provide Opportunities for 21st Century Skill Development?

To present our findings on overall levels of quality in U.S., K-12 wikis, we first display detailed results from our Wiki Quality Instrument measures. We then summarize these results by presenting a taxonomy of four types of educational wikis. Finally, we illustrate this taxonomy with descriptions of typical wikis from each category.

In Table 2, we show the distribution of wiki quality scores at day 14. In the first column, we list values of composite wiki quality scores (from 0-24), and in the second column we present frequency counts of wikis at each composite score value. In the following five columns, we present average WQI subdomain scores within each composite score value.[2] Using the values presented in these five columns, we show how composite wiki quality scores at each value are derived from the five subdomains.

[2] Since wiki quality scores are counts, they possess Poisson distributions, characterized by long upper tails. The use of Poisson regression analysis is more appropriate than OLS linear regression analysis for the modeling of hypothesized relationships between such outcomes and predictors. Similarly, from a strict statistical perspective, the *geometric* – rather than the *arithmetic* – average is a more appropriate summary of the central tendency of Poisson-distributed counts. Typically, in Poisson distributions, the arithmetic mean overestimates the center of the distribution, especially when counts are large. Later in our paper, when we model such relationships and conduct statistical tests, we make use of Poisson regression analysis. However, in the descriptive presentation of Table 2, we made a decision to list the arithmetic – not the geometric—means of the wiki quality subdomain scores. We did this for several reasons. First, we believe readers will find the arithmetic means more intuitive and interpretable. Second, because scores in each of the wiki quality subdomains are typically low, the bias in the arithmetic mean is small or non-existent. Second, whenever *any* of the item scores from which is the geometric mean is constituted are zero, the corresponding geometric mean must also be zero. While anticipated, this would have occurred very frequently in Table 2 had we chosen to display geometric means, concealing what we believe is interesting substantive detail, especially at low wiki quality, which is where the bulk of our sampled wikis fall.

Table 2: *Frequency counts of wikis at each composite wiki quality score value, and average Wiki Quality Instrument subdomain scores within each composite wiki quality score value (n=255).*

Composite WQI Score (24 items)	*n*	**Information Consumption** (2 items)	**Student Participation** (4 items)	**Expert Thinking** (5 items)	**New Media Literacy** (6 items)	**Complex Communication** (7 items)
0	102	0.0	0.0	0.0	0.0	0.0
1	49	1.0	0.0	0.0	0.0	0.0
2	37	1.9	0.1	0.0	0.0	0.0
3	6	0.8	1.0	0.0	0.8	0.3
4	4	0.0	2.5	0.5	1.0	0.0
5	4	0.3	2.5	1.3	1.0	0.0
6	5	0.0	3.0	1.2	1.8	0.0
7	7	0.7	2.6	1.6	1.7	0.4
8	5	0.4	2.6	2.4	2.6	0.0
9	9	0.3	3.1	2.6	3.0	0.0
10	7	0.4	3.0	2.9	3.4	0.3
11	11	0.3	3.1	2.7	4.5	0.5
12	3	1.3	3.3	3.7	3.0	0.7
13	3	2.0	3.7	2.3	4.7	0.3
14	0					
15	0					
16	0					
17	2	1.0	3.5	3.5	3.5	5.5
18	0					
19	1	2.0	3.0	3.0	4.0	7.0
20	0					
21	0					
22	0					
23	0					
24	0					

In Table 2, notice that most wikis have a composite wiki quality score of 0, 1, or 2. There is then a long tail of wikis with scores between 3 and 13, and finally the three highest scoring wikis with scores of 17 and 19. From our analysis of this table (and from analyses not presented here of wiki quality scores at day 30, 60, 100 and 400), we developed a taxonomy of four types of wikis; 1) Trial wikis, failed wikis and teacher resource sites, 2) Teacher-centered, content delivery devices, 3) Individual student presentations and portfolios with limited collaboration, and 4) Collaborative student presentations and workspaces. In Table 3 we show the distribution of wikis among these four categories. To explore these wiki categories, it is illuminating to discuss representative wikis of each type.

Table 3: *Distribution of composite wiki quality scores on day 14 in wikis created in U.S., K-12 public schools, by Title I eligibility (n=255).*

Wiki Quality Score Range	**Wiki Type**	**Public Schools** (n=255)	**Title I Schools** (n=117)	**Non-Title I Schools** (n=133)
0	Failed wikis, trial wikis, or teacher resource sharing sites without student audience or participation	38%	50%	30%
1-2	Teacher-centered content delivery devices	34%	34%	35%
3-15	Individual student assignment or portfolio, with minimal collaboration	25%	15%	35%
16-24	Collaborative, multimedia assignment or workspace	1%	2%	1%

At day 14, 40% of our wikis received a composite wiki quality score of 0, meaning that students had no identifiable interaction with the wiki at all. A typical wiki in this category would have no changes or content. The front page might contain the automatically-generated text created by PBworks ("Welcome to your new wiki!") or some simple modification ("Welcome to Ms. Jones' World History Wiki"). These were trial balloons, which failed to take off. A small proportion of wikis scoring 0 on the WQI served meaningful teacher purposes, such as sharing links or resources, but did not involve students as audience or participants.

Next, 34% of wikis had composite wiki quality scores of 1 or 2, and these 1 or 2 points came from the Information Consumption sub-scale. These were teacher-created, content-delivery devices with students as receivers of information, not content producers. Many of these wikis provided students with syllabi, class policies, teacher contact information, homework calendars, lists of links to resources for research projects, and other features that might be commonly found on a teacher Web site. Some wikis also had newsletters or updates aimed at parents and families. Some content-delivery wikis were quite basic and updated infrequently, and others appeared to be a central part of teachers' communication routines.

In the next category, 25% of wikis had scores between 3 and 15, and these were primarily individually-created student assignments or portfolios.[3] For instance, five wikis had a composite quality score of 6 on day 14, which means that we identified six

[3] We chose to use the composite wiki quality score of 15 as our cutoff point for this category of Individual Student wikis, rather than 13 or some other value, after analysis of wiki quality scores at days 30, 60, 100, and 400. See the online supplement for further details.

behaviors on the wiki that provide an opportunity for students to develop 21st century skills. All of these wikis have elementary features of an individual portfolio, where students have posted simple material about themselves and their hobbies and created a basic site architecture with links and pages for future material. They have no collaborative behaviors and limited evidence of expert thinking or the use of multimedia features.

Returning to Table 2, notice that two WQI subdomains—Expert Thinking and New Media Literacy—are primarily responsible for score differences among wikis with composite quality scores between 3 and 15. The lowest-scoring wikis were spaces where students completed simple tasks, such as writing a basic introduction about themselves in plain text. In a few cases, students posted comments or questions on a teacher-created wiki without interacting with other students. On the highest-scoring wikis in this category, students published multimedia-infused presentations or portfolios on academic topics requiring information organization and crediting of sources.

This activity, however, rarely involved interaction among students. Only 11 of the 63 wikis within this score range have any form of collaboration, and what we found was quite limited: such as students commenting on each other's work or students posting individually-created content to the same page. This finding coheres with previous research about the difficulty of nurturing collaborative wiki environments. For instance, in the evocatively-titled paper, "I DON'T CARE DO UR OWN PAGE!", Grant (2009) provides a case study of how students' strong individual ownership of text prevents a collaborative ethos from developing in a wiki-using, U.K. classroom.

That said, a handful of wikis do involve richer collaboration among students. In our sample, 1% of wikis score above 15 on the WQI by day 14, and these were collaboratively-created, student assignments and workspaces. For instance, the highest scoring wiki was a group presentation about the philosopher Thomas Hobbes. Students used the wiki to collaboratively plan and then co-construct an intellectually rich, multimedia presentation about the philosopher. Another of the top-scoring wikis was used by a middle school literature circle. While reading a novel, students shared responses to reading questions and commented on each other's answers. As a final project, students collaboratively scripted a "movie trailer" for the book, and then used the wiki to plan a video shoot and share multimedia resources for the final edit of the trailer. In these rare cases, students take full advantage of the collaborative and technological affordances of wikis.

In summary, most U.S. K-12 wikis provide few opportunities for 21st century skill development. The majority of wikis are abandoned immediately or are teacher-centered, content delivery devices. An important minority of wikis, however, provide multiple opportunities for students to develop 21st century skills. Most of these wikis are individual productions where students publish assignments or curate portfolios. The few highest-quality wikis are collaborative, multimedia presentations and workspaces.

Do Wikis Created in Schools that Serve More Affluent Populations Provide More Opportunities for 21st Century Skill Development than Wikis Created in Schools Serving Less Affluent Populations?

We find that wikis created in schools serving more affluent populations have more opportunities for 21st century skill development than wikis created in schools serving less affluent populations. In Table 3, we show the distribution of composite wiki quality scores by the Title I status of the hosting school. Notice that 50% of wikis created in Title I schools have a score of 0, compared to 30% of wikis created in non-Title I schools. Also, 15% of wikis created in Title I schools have scores between 3 and 15—the Individual Student wikis—compared to 35% of wikis created in non-Title I schools.

We can test and quantify these differences using Poisson regression. (We summarize our analysis here, and parameter estimates and fit statistics are presented in the online supplement.) From an unconditional Poisson regression model, we can derive the population expected wiki quality score, which is 2.80. We would expect the typical wiki to show about 3 behaviors that provide opportunities for students to develop 21st century skills. When we add to this model a variable controlling for the percentage of students eligible for FRPL in the wiki's school, we find that wikis created in schools with lower proportions of students eligible for FRPL provide more opportunities for 21st century skill development (β_1=-1.59, p<.001). Consider two prototypical wikis, one created in a school with 10% of students eligible for FRPL and another created in a school with 90% of students eligible for FRPL. We estimate that the prototypical wiki created in the high-SES school would have a day 14 composite wiki quality score of 3.82, compared with a score of 1.07 in the prototypical wiki created in the low-SES school.

Quality and Equity in U.S. K-12 Wiki Learning Environments

We found four primary patterns of classroom wiki usage: (1) wikis that were abandoned without being used or used exclusively among educators (40%), (2) teacher-centered content delivery devices (34%), (3) individual student assignments and portfolios (25%), and (4) collaboratively-created, multimedia student presentations (1%). While a minority of all wikis, the wikis in categories (3) and (4) show promise as learning environments that can prepare students for publishing and collaborative problem-solving in a networked age.

From our analysis of wiki equity, we have documented that wikis created in schools serving more affluent students provide more opportunities for 21st century learning and persist longer than wikis created in schools serving less affluent students. There is a great danger that the promise and potential of free Web 2.0 tools will disproportionately benefit the already-advantaged.

These findings, to some extent, cohere with key insights of education technology research from the turn of the century. In his influential book *Oversold and Underused,* Cuban (2001) analyzed technology use in two Silicon Valley high schools. He found that teachers and students used technology infrequently, that few students had student-centered, technology-rich experiences, and teachers nearly universally used technologies to gain efficiencies in established routines rather than to transform practice. We see these patterns echoed in our own findings, but we note two important differences. First, while most wikis are used to gain efficiencies in teacher dissemination of information, a considerable proportion of wikis do provide students with opportunities to publish their writing in a new media platform and a smaller number of wikis foster collaborative student work. If Cuban found a “tiny percentage” of engaging student uses of technology in his landmark research, we find a small but important collection of wikis that do fulfill the potential of networked technologies to support rich learning.

Moreover, Cuban argued that low levels of technology use did not justify the high costs of building school technology infrastructure. With free Web 2.0 tools in an age of ubiquitously wired classrooms, the cost/benefit consideration is different. Many wikis are abandoned soon after creation, but the only cost of such failures is the time of the creator. Teachers who use wikis as content delivery devices have gained efficiencies in typical routines with very low marginal cost. In Cuban's study, disappointing uses of technology came at great financial cost. In our study, disappointing uses of technology come at minimal cost, and they are accompanied by an important minority of wikis that do offer opportunities for students to develop 21st century skills.

In evaluating our findings about wiki learning environments, we also find it useful to consider research about classroom learning opportunities more broadly. One set of studies in 20 midwestern middle and high schools (Nystrand, Gamoran, & Carbonaro, 1998; Nystrand, Wu, Gamoran, Zeiser, & Long, 2003) found that in social studies and English classes, classroom time was primarily dominated by recitation (approximately 20 minutes per class) and lecture (approximately 7 minutes per class). More open discussion occurred in less than 10% of class periods and lasted, on average, less than 1 minute.[4] If these findings are representative of schools nationally, then wikis appear to offer opportunities for publishing, communication, and collaboration that are rarely available in U.S. classrooms.

---

[4] Research projects such as the Gates Foundation Measures of Effective Teaching Project (*Learning about teaching*, 2010), where researchers have video recorded and are analyzing approximately 20,000 lessons, may provide some answers to these questions in the future, presenting another example of how emerging technologies can foster detailed investigations of learning activity at scale.

While we believe that a nuanced examination of wiki quality highlights the potential for Web 2.0 tools to support 21st century learning, we remain concerned about the persistence of Attewell's second digital divide of usage. In our representative wiki sample, we observed patterns that Attewell found in his small ethnographic samples: wikis created in schools serving high-income families provide more opportunities for 21st century skill development than those created in schools serving less advantaged students. Moreover, inequities within schools may be as serious as inequities between schools. While interviewing wiki-using teachers to develop the WQI, several teachers informed us that they used wikis more often or more extensively with their higher-tracked students, who are disproportionately wealthier than their peers. Education technology development and research that does not specifically account for these inequities is likely to exacerbate the second digital divide.

A New Direction for Education Research

One contribution of this research is to provide a detailed portrait of wiki usage in U.S., K-12 settings, with particular attention to how wikis support 21st century skill development and potentially exacerbate digital divides. In crafting this portrait from wiki edit histories, we hope that a second contribution is to present an application for new data sources generated by online learning environments. Without leaving our offices, we made observations from continuously-recorded, student-teacher interactions occurring across the U.S., and from these observations we characterized wiki activity both in depth and at scale.

We suggest several avenues for leveraging these new data. Future research could produce additional large-scale scans examining different wiki providers, Web 2.0 tools, outcome measures, or countries. With this kind of research, researchers could better contextualize the ethnographic and design research that constitutes the core of education technology scholarship. In addition, broad patterns from quantitative content analysis can suggest new avenues for qualitative investigation. For instance, we are intrigued and concerned by our finding that the percentage of wikis created in Title I schools that fail on the first day is twice as high as the percentage of early failures among wikis created in non-Title I schools. Our content analysis will not unravel this puzzle, but teacher interviews and ethnographic approaches could.

Data from online learning environments also has great potential for student assessment (Clarke-Midura & Dede, 2010). In our research, we characterized learning opportunities in wiki communities, and the next logical step would be to use similar analytic methods to track individual student learning. There are no multiple choice tests that will effectively evaluate students' abilities to solve ill-structured problems or collaborate with peers. Online learning environments, however, collect continuous data about students' performance on such tasks. These data could enable the development of time-efficient, valid assessments of higher-order thinking skills. We are optimistic that the earliest forays into this field might rival the efficacy of our current testing systems. For instance, we hypothesize that the number of words that a student writes in secondary school—tracked online—would be a better predictor of college persistence than scores from any contemporary standardized writing assessment. If true, then rather than developing measures of 21st century skills by devising ever more time-consuming testing regimes (Tamayo & Institute, 2010), researchers and policymakers should explore strategies for using real-time, online data sources to measure learning as learners go about their daily activities.

The challenges of realizing the potential of Web 2.0 tools in education are considerable. However, these challenges are paired with new research and assessment opportunities enabled by emerging online-learning platforms. While our research has only touched the surface of these new opportunities, we believe that the analysis of large-scale datasets from online learning environments is one of the most exciting new frontiers of educational research.

References

Alexa. (2010). *Top sites within categories: Wiki farms.* Retrieved May 5, 2010, from http://www.alexa.com/topsites/category/Top/Computers/Software/Groupware/Wiki/Wiki_Farms

Attewell, P. (2001). Comment: The first and second digital divides. *Sociology of Education, 74*(3), 252-259.

Attewell, P. (2003). Beyond the digital divide. In P. Attewell, & N. Seel (Eds.), *Disadvantaged teens and computer technologies* (pp. 15-34). Munster, Germany: Waxmann.

Attewell, P., & Battle, J. (1999). Home computers and school performance. *The Information Society, 15*(1).

boyd, d. (2008). *Taken out of context: American teen sociality in networked contexts.* (Unpublished PhD). University of California, Berkeley, Berekeley.

Chiu, M. M. (2008). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. *Journal of the Learning Sciences*, *17*(3), 415.

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education, 42*(3), 309-328.

Cortez, C., Nussbaum, M., Woywood, G., & Aravena, R. (2009). Learning to collaborate by collaborating: A face-to-face collaborative activity for measuring and learning basics about teamwork. *Journal of Computer Assisted Learning, 25*(2), 126-142. doi:10.1111/j.1365-2729.2008.00298.x

Cuban, L. (2001). *Oversold and underused : Computers in the classroom*. Cambridge, Mass.: Harvard University Press.

Dede, C. (2010). Comparing frameworks for 21st century skills. In J. A. Bellanca, & R. S. Brandt (Eds.), *21st century skills : Rethinking how students learn* (pp. 51-76). Bloomington, IN: Solution Tree Press.

Grant, L. (2009). ‘I DON’T CARE DO UR OWN PAGE!’ A case study of using wikis for collaborative work in a UK secondary school. *Learning, Media and Technology, 34*(2), 105.

Gray, L., Thomas, N. T., & Lewis, L. (2010). *Teachers' use of educational technology in U.S. public schools: 2009.* No. NCES 2010040). Washington, DC: National Center for Education Statistics.

Jenkins, H. (2009). *Confronting the challenges of participatory culture : Media education for the 21st century*. Cambridge, MA: The MIT Press.

Jeong, A. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussions. *American Journal of Distance Education, 17*(1), 25-43.

Larusson, J., & Alterman, R. (2009). Wikis to support the "collaborative" part of collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 4*(4), 371-402.

Learning about teaching - initial findings from the measures of effective teaching project. (2010). Seattle, WA: Bill and Melinda Gates Foundation.

Levy, F., & Murnane, R. J. (2004). *The new division of labor : How computers are creating the next job market*. Princeton, N.J.: Princeton University Press.

Li, C., & Bernoff, J. (2008). *Groundswell : Winning in a world transformed by social technologies*. Boston, Mass.: Harvard Business Press.

Moskaliuk, J., Kimmerle, J., & Cress, U. (2009). Wiki-supported learning and knowledge building: Effects of incongruity between knowledge and information. *Journal of Computer Assisted Learning, 25*(6), 549-561. doi:10.1111/j.1365-2729.2009.00331.x

National Center for Education Statistics. (2007-2008). *Public Elementary/Secondary school universe survey data* [Data File]. Retrieved from Retrieved from http://nces.ed.gov/ccd/

Nystrand, M., Gamoran, A., & Carbonaro, W. (1998). *Towards and ecology of learning: The case of classroom discourse and its effects on writing in high school english and social studies.* No. 11001). Albany, NY: National Research Center on English Learning & Achievement.

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes, 35*(2), 135-198.

O'Reilly, T. (September 30, 2005). *What is web 2.0: Design patterns and business models for the next generation of software.* Retrieved April 20, 2008, from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Oriogun, P., Ravenscroft, A., & Cook, J. (2005). Validating an approach to examining cognitive engagement within online groups. *American Journal of Distance Education, 19*(4), 197-214.

Papert, S. (1980). *Mindstorms : Children, computers, and powerful ideas*. New York: Basic Books.

Reich, J., Murnane, R. J., & Willett, J.B. (2010). *The wiki quality instrument: A tool to evaluate the degree to which teachers use wiki learning environments to promote deeper learning.* Unpublished manuscript.

Reich, J. (2008). *Reworking the web, reworking the world: How web 2.0 is changing our society.* No. Beyond Current Horizons Review paper, Challenge #2). Bristol: Futurelab.

Reich, J., & Daccord, T. (2008). *Best ideas for teaching with technology : A practical guide for teachers, by teachers*. Armonk, N.Y.: M.E. Sharpe.

Richardson, W. (2008). *Blogs, wikis, podcasts, and other powerful web tools for classrooms* (2nd ed.). Thousand Oaks, Calif.: Corwin Press.

Scardamalia, M., & Bereiter, C. (1993). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*(3, Computer Support for Collaborative Learning), 265-283.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis : Modeling change and event occurrence*. Oxford ; New York: Oxford University Press.

Tamayo, J. R., Jr., & Institute, A. (2010). *Assessment 2.0: "next-generation" comprehensive assessment systems. an analysis of proposals by the partnership for the assessment of readiness for college and careers and SMARTER balanced assessment consortium.*Aspen Institute.

Trentin, G. (2009). Using a wiki to evaluate individual contribution to a collaborative learning project. *Journal of Computer Assisted Learning, 25*(1), 43-55. doi:10.1111/j.1365-2729.2008.00276.x

Turkle, S. (2011). *Alone together : Why we expect more from technology and less from each other*. New York: Basic Books.

Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education, 34*(1), 179-225. doi:10.3102/0091732X09349791

Wenglinsky, H. (1998). *Does it compute? the relationship between education technology and student achievement in mathematics.* Princeton, NJ: Educational Testing Services.

Wiske, M. S., Franz, K. R., & Breit, L. (2005). *Teaching for understanding with technology* (1st ed.). San Francisco, CA: John Wiley & Sons.

Zuckerman, E. (2010). International reporting in the age of participatory media. *Daedalus, 139*(2), 66-75. doi:10.1162/daed.2010.139.2.66

The State of Wiki Usage in U.S., K-12 Schools:

Supplemental Materials on Research Design

Given the novelty of the dataset and methods used in our study, readers might find some benefit in a more extensive discussion of our research design than allowed for in the published article presented in the first chapter. In the sections below, we provide additional details on our dataset, samples, instrument, procedures, measures, and data-analytic strategies. In order to prevent the need for readers to move back and forth from the previous chapter to these supplemental materials, we have repeated some information from the body of the first chapter here.

Dataset

PBworks.com is a wiki-hosting service that allows educators and students to set up free wiki workspaces, and it ranks among the top four most-visited sites providing free wikis (Alexa, 2010). From this company, we obtained longitudinal usage data on all 179,581 publicly-viewable, education-related wikis that had been created between the founding of the company in June 2005 through August 2008. These are wikis whose creators designated them as for “Education” during the creation process, as opposed to “Business” or “Personal.”

Each of these 179,581 wikis represents a discrete subdomain on PBworks. The unit of analysis in our study is therefore the wiki subdomain. Hereafter, when we refer to a “wiki” in our dataset, we refer to a publicly-viewable, education-related wiki subdomain hosted by PBworks.

We have both a set of usage statistics on each of these 179,581 wikis and the capacity to examine closely the content of each wiki. We can examine the present state of any wiki, and we can also access every version of every page ever saved during the lifetime of the wiki. We should add one caveat to this statement. Wikis users do have the capacity to delete wiki pages, and if they do so, we cannot investigate the historical records of those pages. In our analysis of wikis, many of which we coded at least four times over the course more than a year, we observed this very infrequently. Most users simply choose to remove pages from the navigational structure, so they become functionally "invisible," rather than delete them entirely. For our analysis of "wiki-level" characteristics, we believe that these infrequent deletions do not affect our findings, but we mention the phenomena as a consideration for researchers pursuing more fine-grained research in wiki learning environments.

In this study, we worked with the entire population of publicly-viewable, education-related PBworks wikis, without restricting our population based on number of edits, number of days of activity, or any other criteria. This preserves our capacity to compare more and less successful wiki learning communities. In particular, retaining wikis with very short lifetimes, few page edits, and no opportunities for student learning allows us to evaluate the full distribution of wiki quality and to compare rates of growth in these characteristics between wikis created in low- and high-SES schools. One common, but we believe misguided, practice used in the study of online environments is to define the population of interest using an outcome measure or proxy for the outcome measure. Selecting a population on the outcome of interest causes undesirable biases in population estimates (Murnane & Willett, 2011). More importantly, eliminating these cases from study restricts the ability of researchers to identify differences between projects with desirable and undesirable characteristics.

In addition, we supplemented our wiki-level data by merging aggregate-level school data from the *Common Core of Data* (Sable & Plotts, 2009).

Sample

For the analyses presented here, we drew a 1% random sample of 1,799 wikis from the population of 179,581 education-related wikis made available to us by PBworks. Unlike many circumstances where population-level statistics are unknowable, here we have some capacity to evaluate the effectiveness of our random sampling. Although we report on wikis drawn from our 1% sample in the paper, we do have access to a number of descriptive statistics concerning the entire population of 179,851 wikis.

In Table 1, we present a table containing descriptive statistics to compare our 1% sample of 1,799 wikis with the rest of the entire population. At the time of initial data collection, our sample and its parent population had no differences in the mean values of page saves per wiki, of wiki lifetimes in seconds, and of the time difference between the location of the wiki creators' IP address and Greenwich Mean Time. The sample and population did not differ with regards to the proportion of wikis with premium wiki features and the proportion of wikis started in January through November. Our population and sample are so large, we can detect very small but statistically significant differences in the proportion of wikis started in December and the proportion of wikis that adopted version 2.0 of the PBworks graphical user interface. These differences are quite minimal: 4.7% of wikis in the population started in December, versus 3.7% percent of wikis in our sample, and 28% of wikis in the population had adopted version 2.0 of the PBworks user interface by August of 2008, compared to 25.5% of wikis in the sample. So by multiple observable measures, our sample can be confirmed to be representative of our population, and for those two measures where they differ, they differ by modest amounts. We believe this provides strong evidence of the success of our randomization.

Table 1*: Descriptive statistics of wikis in our 1% sample (n=1,799) and all other wikis in the population (n=178,052).*

Variable	**Group**	**N**	**Mean**	**Std Dev**	**Std Err**	**Min**	**Max**	**DF**	**t statistic**	**Pr > |t|**
Page Saves	*All-but- sample*	178052	31.4066	174.7	0.4139	0	19756			
	Sample	1799	32.2151	205.4	4.8437	0	7235	179849	-0.19	0.8454
Time Alive in	*All but- sample*	178052	3789316	9307781	22058.3	0	1.02E+08			
	Sample	1799	3766087	8706723	205276	0	66443100	179849	0.11	0.9161
Timeshift from GMT	*All but- sample*	178043	-3.0608	4.371	0.0104	-12	19			
	Sample	1799	-3.1982	4.2585	0.1004	-12	13	179840	1.33	0.1847
Premium Features #1	*All but- sample*	178052	0.0049	0.0696	0.000165	0	1			
	Sample	1799	0.0017	0.0408	0.000962	0	1	179849	1.95	0.0518
Premium Features #2	*Allbut - sample*	178052	0.0056	0.0749	0.000177	0	1			
	Sample	1799	0.0067	0.0814	0.00192	0	1	179849	-0.58	0.5613
Premium Features #3	*All but- sample*	178052	0.0018	0.0428	0.000101	0	1			
	Sample	1799	0.0011	0.0333	0.000786	0	1	179849	0.71	0.4769
GUI Version 2	*All but- sample*	178052	0.2800	0.449	0.00106	0	1			
	Sample	1799	0.2557	0.4364	0.0103	0	1	179849	2.29	0.0223
January Start	*All but- sample*	178052	0.0744	0.2625	0.000622	0	1			
	Sample	1799	0.0773	0.2671	0.0063	0	1	179849	-0.45	0.6496
June Start	*Allbut - sample*	178052	0.0889	0.2846	0.000675	0	1			
	Sample	1799	0.0823	0.2748	0.00648	0	1	179849	0.99	0.3236
Dec Start	*All but- sample*	178052	0.0472	0.212	0.000502	0	1			
	Sample	1799	0.0367	0.188	0.00443	0	1	179849	2.09	0.0369

From the 1% sample of 1,799 wikis, we identified 406 wikis created in U.S., K-12 schools for further study. We disqualified 18 wikis that were set to be privately viewable (removed from public view) during the observational period, 502 wikis that were either deleted or never changed (which unfortunately are collapsed in one category—we believe that the vast majority of wikis in this category were never changed), 448 wikis that were used exclusively outside the U.S., and 425 U.S. wikis that were not identifiably from K-12 settings (most of which were from higher education).

Of the remaining 406 wikis, 43 were created in independent schools or home-schooling organizations serving K-12 students. Two were created in public libraries, three in university settings serving K-12 students (e.g. a summer school) and 36 were from unknown sources. The remaining 322 were created within the U.S. public school system. We could identify particular schools for 255 of these wikis. For 54 wikis, we could identify only a district affiliation. Some of these had no particular school affiliation because the served district teachers broadly, and for others we could identify the hosting district with certainty but we could not identify a particular school with certainty. Thirteen wikis were hosted by district collaborative organizations like BOCES in New York or Area Education Associations in Iowa.

Since we can match only school-level demographic data on wikis from specific public schools, we limit our attention to the subpopulation of 255 wikis created in identifiable public schools, rather than focusing on our full sample of 406 K-12 wikis.

That said, the distribution of wiki quality in the full sample does not differ from the corresponding distribution in the public school sample.[1]

For the purposes of the survival analyses reported in the paper, we evaluated all wikis as of March 28, 2009. Since we drew our sample of wikis based on all of the wikis created through August 2008, this meant that every wiki had at least seven months to develop. Thus, in our survival analyses, our study had 1187 days of wiki-accrual time (June 2005 through August 2008) and 209 days of follow-up time (August 2008 through March 2009). With our sample of 255 public-school wikis, in our survival analysis we had a statistical power of .90 to detect effects of moderate size, such as 100% differences in median lifetimes between sub-populations. (Singer & Willett, 2003a).

Sample Limitations

We note two limitations of our sample. First, we have access only to PBworks wikis, raising questions as to whether PBworks wikis are representative of freely available wikis hosted by other providers. The only major comparable alternative host is Wikispaces.com. PBworks and Wikispaces trade places from week to week as ranked 3 and 4 on the Alexa rankings site for wiki hosting services (Wikia and WetPaint, which do not have a significant share of educational wikis, rank 1 and 2 typically). There is considerable anecdotal evidence that PBworks and Wikispaces serve very similar audiences. Online threaded discussions and blog posts by practicing educators and school

[1] To test the hypothesis that the distribution of wiki quality among our 406 U.S. K-12 wikis does not differ based on public school status, we fit a Poisson regression model predicting ln(composite wiki quality) conditional on whether or not the wiki was in our subsample of 255 identifiable public school wikis. We found that wiki quality does not differ by whether or not a wiki was in the public school subsample (β_1=-.04, p=.51)

technology staff addressing “best tech tools for education” and “best wikis in education” typically discuss both providers, if they discuss one. The functional differences between the two sites are minor. As we designed our *Wiki Quality Instrument* (below), we conducted 68 interviews with teachers who had used PBworks, Wikispaces, and both platforms. Typically, teachers explained their choice of platform as a function of which one they happened to encounter first online or with a colleague, rather than because one offered a distinctive affordance.

On the social network Classroom 2.0, there has been an active discussion about the affordances of both platforms (http://www.classroom20.com/forum/topics/wikispaces-vs-pbworks). No consensus has emerged in this discussion about functional differences between the platforms or the superiority of one or the other for particular purposes or populations. Representatives of both companies weighed into the discussion and neither claimed to offer a superior product. Indeed, one Wikispaces employee wrote:

> James from Wikispaces here. I'll answer your questions in detail in a moment, but first a few thoughts about choosing a wiki for your classroom. We're big believers at Wikispaces in using the solution that's most comfortable for you and your students, especially when you're just getting started. We spend a lot of effort making Wikispaces very easy to jump into with limited time and setup, and our focus on educational use nearly defines our company. **If you're already happily using PB Works, there might not be a compelling reason to switch** -- you've already done the hard work of getting up and running. (emphasis added)

So, to date, we have no reason to believe that PBworks and Wikispaces wikis would differ significantly from each other in regards to teacher and student usage. Of course, the best way to settle the issue would be to conduct a separate, full-replication study using data from Wikispaces.

The second limitation of our sample is that it does not include privately viewable wikis. To begin with, publicly-viewable wikis represent 70% of the wikis created on PBworks from 2005-2008, so even if our findings are only generalizable to the population of public wikis, they are generalizable to the majority of wikis. It is not clear to us whether we would expect privately viewable wikis to be used differently. While many might assume that most wikis with student activity would be kept private, there is extensive evidence of publicly-viewable student activity in our analytic sample. In our teacher interviews, we found that teachers' decisions as to whether to make a wiki privately viewable or not tended to be driven more by school and district culture than by the particular activities planned by wiki-using educators. In some schools, greater attention to student privacy was a strong norm and online learning environments were expected to be password protected. In other places, teachers had more leeway to decide.

Resolving this dilemma is quite difficult, since by their very nature it is not possible to obtain a listing of privately viewable wikis from which to draw a representative sample. Any study of privately viewable wikis will necessarily be limited by selection bias, since researchers can only view these wikis by permission. Currently, we are conducting a study involving an automated survey to all PBworks wiki creators where we solicit the URL of both privately and publicly viewable wikis.

Instrument

To measure wiki quality, we used the Wiki Quality Instrument (WQI) to code the content of individual wikis and assess the opportunities they provided for students to develop 21st century skills. We developed and piloted the WQI over an 18-month period using a rigorous design process. This process included a year of ethnographic fieldwork in wiki-using classrooms to assess how teachers and students themselves evaluate wiki quality, a detailed literature review of studies that have used content analysis to evaluate online learning environments, and several rounds of pilot testing and revision.

The WQI has two sections. In the first section, we gather demographic information about the wikis, such as a wiki's academic subject area, hosting institution, creator, participants, and audience. Items in the second section measure wiki quality in terms of five subdomains: (a) Information Consumption (2 items) (b) Student Participation (4 items), (c) Complex Communication (7 items), (d) Expert Thinking (5 items), and (e) New Media Literacy (6 items).

In each subdomain, coders assess whether students participate in activities that provide opportunities to develop 21st century skills: activities such as co-creating a shared page, reflecting on a work product or process, copyediting another person's work, or embedding multimedia frames into a page. A listing of the 24 questions for these items is presented as Table 1 in the main paper, and is reproduced as Table 2 below. To obtain our composite wiki-quality scores for each wiki, we sum the values of these 24 dichotomously coded items to generate a measure ranging from 0-24.

Table 2: *Summary questions of coding categories used in the Wiki Quality Instrument*

Information Consumption	
Course Materials	Do students come to the wiki to access academic materials?
Information Gateway	Do students come to the wiki to access links to other Web sites?
Student Participation	
Contribution	Does at least on student contribute, in any form, to the wiki?
Individual Pages	Does at least one student own their own page on the wiki?
Shared Pages	Does at least one pair (or group) of students own their own wiki page?
Ownership	Do student(s) serve as primary facilitator and content creator of the wiki?
Expert Thinking	
Academic Knowledge	Does at least one student complete a task requiring academic knowledge (as opposed to simply writing about hobbies or one's family)?
Information Organization	Does at least one student complete a task requiring information organization, rather than routine information retrieval?
Metacognition	Does at least on student reflect on his/her work product or process?
Crediting	Does at least one student credit their sources of his/her work?
Teacher Feedback	Do teachers provide feedback on student work?
Complex Communication	
Concatenation	Do multiple students add discrete sections of text to the same page?
Copyediting	Does at least one student copyedit text created by another student?
Co-Construction	Does at least one student substantively edit text created by another student?
Commenting	Does at least one student comment upon another students work on the wiki?
Discussion	Do students respond to each others' comments for at least four conversational turns?
Scheduling	Do students schedule meetings or tasks?
Planning	Do students plan for future work?
New Media Literacy	
Formatting	Does at least one student use formatting elements beyond plain text?
Links	Does at least one student post a link to another page or document?
Hyperlinks	Does at least one student create links rendered as simple text or images?
Images	Does at least one student embed an image into a page?
Uploads	Does at least one student upload a document?
Multimedia	Does at least one student embed a multimedia element into a page?

In Table 3, we present estimated Cronbach's coefficient alphas and corresponding inter-rater agreement averages, for each subdomain. Cronbach's alpha estimates range from .68 (in the subdomain with two items) to .86, and average inter-rater agreement in each subdomain ranges from .78 to .96. These summaries demonstrate suitable inter-rater agreement within each section, and suitable cohesion within the subdomains. Inter-rater agreement is discussed in more detail in the next section on procedures.

Table 3*: Cronbach's Coefficient Alphas and average inter-rate agreement for the five subdomains of the Wiki Quality Instrument.*

	Information Consumption	Participation	Expert Thinking	Complex Communication	New Media Literacy
Cronbach's Alpha	.68	.86	.80	.85	.86
Average Inter-rater Agreement	.78	.90	.93	.96	.91

Procedures

The wikis in our sample are extremely diverse. They are used with elementary schools through high schools, in nearly every subject area imaginable, and for a wide variety of educational purposes. They range in size and complexity from a single page with no revisions to wikis with hundreds of pages revised thousands of times. Accurately characterizing the activity on wikis is very challenging work. In this section, we present our strategies to meet these challenges.

The wikis in our analytic sample were coded using a multistage process, and each process was piloted and revised several times before attempting a final analytic sweep. To identify basic demographic features of each wiki, two trained research assistants read each sampled wiki to code the wiki's hosting school, subject area, and grade level. From preliminary analysis, we developed 15 non-exclusive subject-area classifications (Contained Elementary, English/Language Arts, Math, Science, English as a Second Language, Social Studies, Computer Science, Modern Foreign Language, Classic Foreign Language, Visual and Performing Arts, Business, Library, Education, Physical Education and Health, and Other) and three non-exclusive grade level classifications (K-5, 6-8, 9-12). Because of the importance of identifying hosting schools for matching school-level socio-economic data, a third research assistant reviewed all wikis again to confirm hosting-school identifications. This school-identification work involves quite a bit of "detective" work. Some wikis were clearly labeled as being used in a particular school. Other wikis were identified through email addresses on the wiki, words or initials found in the subdomain URL, or names of teachers that could be identified in lists of school personnel and corroborated through other evidence.

The first author then resolved coding disagreements among raters. Inter-rater agreement estimates demonstrate that our coding achieved acceptable levels of reliability. For our subject categories, raw inter-rater agreement averaged .96 and ranged from .90 to .99. For our grade-level analysis, raw inter-rater agreement averaged .83 and ranged from .78 to .86.

With these demographic features established, we then coded wikis to determine the users of the wiki and the opportunities for students to develop 21st century skills measured by the Wiki Quality Instrument. Before beginning the coding process, we established an extensive training regime. We developed a training set of coded wikis, and research assistants had to reach 85% agreement with the first author across all quality categories of the WQI, and have a final average composite wiki-quality score that fell within 1.5 points of the first author's rating, before being allowed to begin coding new wikis. To maintain a close alignment of scores, research assistants participated in weekly meetings to discuss wikis and quality categories that were particularly difficult to code.

To evaluate quality longitudinally in our sample of wikis, research assistants used the WQI to measure quality on six occasions: at 7, 14, 30, 60, 100, and 400 days from wiki creation. To facilitate the administration of the WQI on the historical record of wikis, we used the Wiki Coding Tool. This tool is a Web interface that draws on the PBworks' data warehouse and permits a coder to examine the appearance of a PBworks wiki at any particular day in the wiki's development. Because the entire historical record of every edit to every page of every wiki is stored by PBworks, our Wiki Coding Tool is a "time machine" for assessing wiki usage. Using it, we can sample the appearance of wiki pages on any day of a wiki's existence and review all changes and edits between occasions of measurement.

Each wikis was coded by two raters at each designated occasion of measurement as long as the wiki continued to change. Thus if a wiki's final change was on day 25, it was coded on days 7, 14, and 30, but no further. On each occasion of measurement, the two raters evaluated every revision to every page, all page comments, and all documents uploaded to the wiki up to that time period. On small wikis with only one page, this might take only a few minutes. On the largest, most complex wikis on their 400^{th} day, this process can take several hours.

Since the Wiki Quality Instrument measures opportunities that a wiki provides for students to develop 21^{st} century skills, on each occasion of measurement we evaluated all of the activity occurring on the wiki up through that occasion of measurement. Thus, an evaluation on day 7 included an assessment of all activity from creation through day 7, and an evaluation on day 14 includes all activity from creation through day 14. Wiki-quality scores, therefore, are monotonic.

In our first round of wiki-quality coding, research assistants first identified the creator and participants in the wiki. The creator of a wiki is defined as the primary content creator or facilitator of the wiki. This is usually the person who actually submits the request to PBworks.com to generate the unique subdomain, but occasionally an educator will create blank wiki subdomains for students, who then go on to populate the wiki with content. Research assistants coded wiki creators in three exclusive categories: student, educator, and other. Participants are defined as people who contribute to a wiki who are not the creator. Research assistants then coded wiki participants in three non-exclusive categories: student, educator, and other. After identifying the users of a wiki, the researchers coded their responses to the 24 dichotomous categories of the WQI.

During this first round, inter-rater agreement appeared to be generally high and final composite wiki-quality scores usually differed no more than two points between the two raters, but one issue emerged. Occasionally, two raters would disagree about the creator of a wiki, one believing the creator was a student and another believing the creator to be an educator or to be not knowable. Since most categories in the WQI evaluate student behaviors, this could lead to large differences in scores. This led to the addition of a review protocol in our coding process. After two raters submitted their ratings for a given wiki, if they disagreed about the creator or participants of a wiki, if they had differences in quality scores greater than 3, or if one rater scored a wiki as 0 and the second rater gave the wiki a positive score, then the wiki was flagged for review. The two raters would then meet to discuss only their user categories—not their quality scores—and to explain their evidence for believing the creator was a student, educator, or unknowable. They then revised their user and quality codes individually based on the discussion and resubmitted them. This process inflates our agreement on user categories but resolved some of the major discrepancies in wiki-quality coding. As a final step, a third, senior research assistant resolved all remaining disagreements. In our six user categories, inter-rater agreement averaged .95 and ranged from .91 through .99.

In terms of quality-code agreement, our research team coded 406 U.S., K-12 wikis at 1219 time points, an average of 3 occasions of measurement for each wiki. In Table 3, we present the average inter-rater agreement for the five subdomains of the WQI over all occasions of measurement, which ranged from .78 through .96. Average inter-rater agreement across all 24 items and over all occasions of measurement was .92. It should be noted that estimating these inter-rater agreements over all occasions of measurement produces a downward bias in agreement scores. If two raters disagree about the presence of an opportunity on day 7, and then continue to disagree through day 400, then one fundamental disagreement is repeated six times in our analysis.

We also estimated mean composite wiki-quality scores and mean differences in quality scores between raters. Keep in mind that since our scores are likely Poisson distributed, the mean is an inflated summary of the center or location of the distribution, since we expect our “count”-like score distributions to have long upper tails. The mean quality score of both raters over all 1219 measurements was 3.59, out of a possible score of 24. The mean difference in scores between the two raters was 1.46. In 47% (576) of measurements, the two raters were in perfect agreement. In 25% (309) of measurements, there was a score difference of exactly 1, a difference of 2 in 11% (129) of measurements, a difference of 3 in 6% (79) of measurements, and a difference of more than 3 in 10% (125) of measurements, including 19 measurements where scores differed by more than 10 points—these representing occasions where the two initial raters persisted in disagreeing about the identity of the person(s) editing the wiki. Of course, all rater disagreements were all reconciled by a third senior rater.

As these measures indicate, we had considerable success in providing a reliable, quantitative evaluation of extraordinarily diverse content. Our estimates of agreement may not be as high as what can be achieved by researchers examining a narrower set of learning environments and evaluating a narrower domain of behavior, but given the complexity of our sample and of our measures, we believe that we have generated adequate measures of wiki quality for use in our national assessment of wiki use.

One other limitation of our quality coding is worth highlighting. When administering the WQI, we coded seven different forms of collaborative behavior, listed in Table 2. In evaluating students’ collaborative behaviors, we were dependent upon students logging in with their own user ID or leaving bylines associated with their contributions (e.g. “Irish History, by Jane McDonnell”). In many cases, students observed these norms, and we were able to measure collaborative activity with precision. We know from our observations, however, sometimes students do not log in under a unique ID and sometimes multiple students work on a page while logged in under one ID, perhaps while sitting next to each other and sharing a computer in a school lab. We cannot credit collaborative activity that we cannot identify affirmatively. Therefore, it is possible that we have underestimated collaborative activity within our sample of wikis. This was a topic of discussion on several occasions in our weekly team meetings, and the consensus of our team was that over the hundreds of wiki we evaluated, raters felt that there were few occasions where they believed they might be under-representing collaboration because of ambiguities in user identity. In a sense, we resolved this issue by clearly defining the collaborative behaviors that we measured as behaviors that happen within the wiki environment. We did not measure dimensions of collaboration happening within classrooms and computer labs, dimensions which are certainly important and worthy of ethnographic study.

We confirmed the validity of our measures by comparing wikis with similar quality scores. For instance, in the main paper, we characterize all five wikis with a quality score of six on their fourteenth day. Each of these wikis serves a similar purpose as the simple foundation of an online, individual student portfolio. The ultimate test of validity would be to conduct studies within specific school contexts that measured student development in 21st century skills as a function of their participation in wiki learning environments, and we hope in future studies to bring the insights from this national perspective on wiki usage to study individual student development with Web 2.0 learning environments.

Measures

As noted in the paper, our analyses of our longitudinal wiki-quality measures revealed that wiki-quality trajectories tended to be nonlinear, with the quality being a logarithmic function of time. Wiki quality typically increased rapidly within the first 14 days after wiki creation and then quality growth leveled. As a result, in our cross-sectional analyses, we chose to use composite wiki-quality scores at day 14 as our outcome measure. Therefore, for these analyses, we record the values of all our measures in a project-level dataset, where each row corresponds to one wiki as evaluated on its 14th day. In Table 4, and below, we briefly define and summarize our measures in three sections: outcome measure, question predictor, and additional predictors.

Table 4*: Definitions of principal variables included in the analyses.*

Category *Sub-category*	**Variable**	**Decision Rule**
Outcome		
Wiki Quality Score	*WQUALITY*	A continuous variable ranging from 0-24 which measures the number of observed behaviors that occur on a wiki that represent opportunities for students to develop 21st century skills such as expert thinking, complex communication, and new media literacy.
Question Predictor		
School-Level SES	*PERFRPL*	A continuous variable ranging from 0.00 to 0.99 representing the the proportion of students in a wiki's hosting school eligible for Free or Reduced Priced Lunches
	TITLEI	A dichotomous variable coded as "1" when a wiki's hosting school is eligible for Title I funding and "0" when a wiki's hosting school is ineligible.
Additional Predictors		
Lifetime in Days	*DAY*	Continuous variable representing the numbers of days between a wikis creation and the final wiki edit

Outcome Measure

WQUALITY is a continuous variable, ranging from 0-24, that summarizes the degree to which a wiki provides opportunities for students to develop 21st century skills such as expert thinking, complex communication, and new media literacy. We obtained this measure by totaling the coded responses to the 24 questions of the Wiki Quality Instrument after analyzing all changes on the wiki through the 14th day after wiki creation. Since these measures are the sum of counts, our outcome measure has a Poisson, rather than a normal, distribution.

Question Predictors

In addressing our second research question concerning inequities in opportunities with wikis created in schools serving different populations, our question predictor was *PERFRPL*. *PERFRPL* is a continuous variable, ranging from 0.00 to 0.99 that measures the proportion of students in a school eligible for the Free or Reduced-Priced Lunch program. We obtained these data from the Common Core of Data.

In our survival and contingency-table analyses, we also used a school's Title I eligibility as an indicator of school-level socioeconomic status (SES). Schools are eligible for Title I funding if more than 40% of students come from families living below the federal poverty line (Sable & Plotts, 2009). *TITLEI* is a dichotomous variable coded "1" for wikis created in schools eligible for Title I funding, and coded "0" in public schools not eligible for Title I funding.

Additional Predictors

We also used measures of time to conduct the survival analysis featured in the section on Patterns of Wiki Use in the main paper. In order to use measures of wiki lifetimes to evaluate wiki usage, we applied a biological metaphor, the lifecycle, to a socio-technical community. We measured wiki lifecycles in days from wiki “birth” through wiki “death.” The birth of a wiki occurs at a distinct, measurable moment when a user generates a new subdomain on a wiki hosting network. Day 0 is the period of time from wiki creation until 11:59PM on the same day. Day 1 concludes at 11:59PM on the day after wiki creation, and so on.

Designating the moment of death, or failure, of a wiki is more subjective than is identifying its creation, since wikis can always be returned to, changed and edited, even after years of inactivity. Nonetheless, we could identify precisely the last moment when a wiki was changed (through a page edit or new page creation), after waiting a sufficient time without further activity to ensure that the wiki is not merely dormant. Since the longest break in the U.S. academic year is the three-month summer holiday, we adopted a 90 day period of inactivity as being sufficiently long to designate a wiki as “dead.” In our survival analysis of wikis, *DAY* records the number of days from wiki creation to a final wiki edit, and it ranges from 1 to 914.

Data-Analytic Strategy

Descriptive Statistics

We estimate survivor functions for our entire 1% sample of 1,799 wikis and our subsample of 255 public school wikis. We did this using the Kaplan-Meier approach (Kaplan & Meier, 1958). We also compared the survivor functions of wikis created in Title I eligible and non-Title I eligible schools. To do this, we obtained Kaplan-Meier estimates of the survivor functions separately for wikis from Title I eligible and non-Title I eligible schools in the project-level dataset, and we compared them using a Wilcoxon rank test (Kalbfleisch & Prentice, 1980; Singer & Willett, 2003b). For each survivor function, we estimated the corresponding median lifetime.

RQ#1 To what extent do wikis provide opportunities for 21st century skill development in U.S. public schools?

In the main paper, we used two kinds of analyses to answer this question. First, we presented the distribution of empirical composite wiki-quality scores. From these values, we offered a taxonomy of four wiki types based on the composite wiki-quality scores: wikis with a score of 0, with a score of 1 or 2, with a score of 3 through 15, and with a score greater than 15.

The division lines at composite wiki-quality scores of 0 and 2 are intuitive, and the division point at 15 is more subjective. In addition to evaluating wiki-quality scores at day 14, we also evaluated wiki-quality scores at the final occasion of measurement (through day 400 or the final wiki edit, whichever was earlier). Evaluating wiki-quality scores through this lens has the advantage of seeing each wiki used for the full duration of its lifetime, or at least for 400 days. In the main paper, we chose not to use this approach in favor of presenting day 14 wiki-quality scores, which provide a more consistent comparison. Measures of wiki quality at the time of the final edit were somewhat higher than day 14 wiki-quality scores, and there were several more wikis that score above 10 points on the WQI. We found that among wikis with scores below 15 on their final day, the average Complex Communication subdomain score was low. On wikis with scores of 15 or higher, average Complex Communication subdomain scores were 4 and higher. Hence, we used 15 as the cut point between our categories of Individual Student wikis and Collaborative Student wikis.

To quantify and summarize the distribution of composite wiki-quality scores, we fitted an unconditional Poisson-regression model where the outcome is effectively log(*WQUALITY*) and where composite wiki-quality scores are assumed to have a Poisson distribution. In Table 5, we display a taxonomy of fitted Poisson-regression models. In the unconditional model, the parameter estimate associated with the constant is 1.03 (p<.001). To obtain an estimated population “average” (expected) wiki-quality score, we antilogged this parameter estimate to obtain the value 2.80, published in the main paper.

Table 5: *Taxonomy of fitted Poisson regression models estimating log composite wiki-quality score at day 14 (n=248; FRPL data is missing for 7 wikis).*

	Unconditional	SES
Intercept	1.03***	1.50***
	(.04)	(.06)
Percentage of students eligible for FRPL		-1.59***
		(.17)
-2LL	1680.03	1580.98

Cell contents are parameter estimates and (standard errors). *** p<0.001

Note. Expected composite wiki-quality scores are obtained by $\hat{Y} = e^{\beta_1 X_1} * e^{\beta_2 X_2} ... * e^{\beta_x X_x}$.

RQ#2 Do wikis created in schools that serve more affluent populations provide more opportunities for 21st century skill development than wikis created in schools serving less affluent populations?

To address this question, we added the question predictor *PERFRPL,* which describes the socio-economic level of the school in which the wiki was created, to our Poisson regression model. Thus, we fitted the following model, where for each wiki

$$\log(WQUALITY) = \beta_0 + \beta_1 PERFRPL$$

where *WQUALITY* is assumed to have a Poisson distribution. We obtained fitted values for prototypical wikis by antilogging estimates of $\beta_0 + \beta_1 PERFRPL$. In Table 5, we show that when we fitted this model to our data, we estimated β_0 to be 1.50 (p<0.001) and β_1 to be -1.59 (p<0.001). Therefore, to obtain the estimated wiki-quality score for a prototypical wiki created in a school with 90% of students eligible for FRPL, we take the antilog of 1.50+-1.59*.09, which is 1.07. To obtain the estimated wiki-quality score for a prototypical wiki created in a school with 10% of students eligible for FRPL, we take the antilog of 1.50+-1.58*.01, which is 3.82.

Conclusion

Researchers who have further questions about our methods or who are interested in conducting similar studies are encouraged to contact the first author at bjr795@mail.harvard.edu.

References

Alexa. (2010). *Top sites within categories: Wiki farms.* Retrieved May 5, 2010, from http://www.alexa.com/topsites/category/Top/Computers/Software/Groupware/Wiki/Wiki_Farms

Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457-481.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter : Improving causal inference in educational and social science research*. Oxford ;New York: Oxford University Press.

Sable, J., & Plotts, C. (2009). *Documentation to the NCES common core of data public Elementary/Seconday school universe survey: School year 2007-2008.* No. NCES 2010-302). Washington, D.C.: National Center for Education Statistics.

Singer, J. D., & Willett, J. B. (2003a). Survival analysis. In I. B. Weiner, & D. K. Freedheim (Eds.), *Handbook of psychology* (pp. 555-581). Hoboken, N.J.: John Wiley.

Singer, J. D., & Willett, J. B. (2003b). *Applied longitudinal data analysis : Modeling change and event occurrence*. Oxford ; New York: Oxford University Press.

The Wiki Quality Instrument: An Instrument for Measuring Opportunities to Develop 21st Century Skills in Wiki Learning Environments

Introduction
Part I: The Instrument: Summary, Definitions and Decision Rules
Part II: Content Analysis and Training Guidelines and Protocols
Part III: Developing the WQI
Part IV: Developing the WQI Protocols
Part V: Adaptation Guidelines for Educators
Part VI: Adaptation Guidelines for Researchers

We designed the Wiki Quality Instrument (WQI) to conduct a study examining issues of excellence and equity in the educational use of wikis in U.S. K-12 settings. Our hope is that both the WQI and the process for designing and administering the WQI might prove to be of use to other researchers interested in doing large scale content analyses that take advantage of the massive datasets maintained by online learning environments.

In our research study, we examined wikis that were used in kindergarten through high school classrooms, in many different academic subject areas, and for many different purposes. Our version of the instrument is tailored to study opportunities for 21st century skill development in this particular context. The first publication resulting from that study was “The State of Wiki Usage in U.S., K-12 Schools: Leveraging Web 2.0 Data Warehouses to Assess Quality and Equity in Online Learning Environments,” which appeared in the January 2011 issue of Educational Researcher. Reading this study is the ideal introduction to how the WQI was used to characterize the use of wikis across the U.S. in detail and at scale.

On this website, we provide several resources for using and adapting the Wiki Quality Instrument. For researchers interested in studies similar to our own—studies of wikis in U.S., K-12 settings—we provide our coding manual, our training guidelines, and our evaluation protocols. We expect, however, that most researchers and educators will be interested in slightly different contexts and therefore interested in modifying our instrument and protocols. As a result, we provide two documents describing the development of the WQI and its protocols, and we also make suggestions for adapting the WQI for educational contexts and for other research contexts.

All of the materials on this site are licensed with a Creative Commons CC:BY license, and researchers and educators are encouraged to modify and re-use the materials. We welcome comments and questions, and we would especially appreciate hearing from other researchers using these tools: justin@edtechteacher.org. Our work has been generously supported by the Hewlett Foundation's Open Education Resources initiative.

Introduction

The Wiki Quality Instrument (WQI) is a coding tool for evaluating the degree to which wikis provide opportunities for students to develop 21st century skills such as expert thinking, complex communication, and new media literacy. The tool was developed by Justin Reich, project manager of the Distributed Collaborative Learning Communities project, in consultation with Hunter Gehlbach, Richard Murnane, and John Willett over an 18-month period. The development process included extensive qualitative research in wiki-using classrooms, a literature review of research on 21st century skills and on evaluating skill development in online learning environments, and multiple rounds of pilot testing and iteration.

The main section of the Wiki Quality Instrument consists of 24 items divided into five subdomains. Two items on Information Consumption evaluate ways in which students use the wiki to access academic resources. Four items in the Participation subdomain evaluate the basic ways in which students contribute to the wiki. Five items in the Expert Thinking subdomain evaluate opportunities that students have to organize information, solve academic problems, reflect on their learning, credit sources, and receive feedback from educators. The seven items in the Complex Communication subdomain evaluate the degree to which students communicate and collaborate with other students. The six items in the New Media Literacy category measure the degree to which students use the various technical affordances of the wiki to share content, links, images, and multimedia. The WQI also includes demographic questions about grade level, academic subject area, hosting site, creators, participants and audience. Finally, the WQI includes four overall rating questions to be evaluated subjectively by raters.

We designed the WQI to conduct a series of studies examining the use of wikis in U.S. K-12 settings. We drew random samples of wikis from populations of hundreds of thousands of wiki learning environments. Trained research assistants then used the WQI to assess the edit histories of each sampled wiki at multiple time points in the wiki lifecycle. Thus, we represent wiki quality as a trajectory, a series of linked measures, rather than as a single numerical point. In our research, we both assess the shape of typical wiki-quality trajectories, and we estimate the degree to which covariates such as grade level, academic subject area, school-level socio-economic status and teacher attitudes predict differences in the initial position or rate of change in these wiki-quality trajectories.

On this website, we provide several resources for using and adapting the WQI. We designed the WQI to conduct a study of wikis used in U.S., K-12 settings. Our samples of wikis were used in Kindergarten through high school classrooms, in many different academic subject areas, and for many different purposes. Our version of the instrument is tailored to study opportunities for 21st skill development in this particular context. For researchers interested in similar contexts, we provide our coding manual, our training guidelines, and our evaluation protocols. We expect, however, that most researchers and educators will be interested in slightly different contexts and therefore interested in modifying our instrument and protocols. As a result, we also detail the processes by which we developed the WQI and its associated protocols, and we make suggestions for adapting the WQI for educational contexts and for other research contexts.

All of the materials on this site are licensed with a Creative Commons CC:BY license, and researchers and educators are encouraged to modify and re-use these materials. We welcome comments and questions, and we would especially appreciate hearing from other researchers using these tools. Please use the contact form at the top-right of the page to get in touch with us.

Our work has been generously supported by the Hewlett Foundation's Open Education Resources initiative.

Part I: The Wiki Quality Instrument

Section 1: Summary Questions
Section 2: Definitions
Section 3: Decision Rules

Section 1: Summary Questions

We present these summary questions to introduce new educators and researchers to the WQI. These summary questions were not used by our trained researchers. In piloting the WQI, we estimated that we achieved the most valid assessments and highest inter-rater agreement among trained coders when we used short, declarative sentences as decision rules. However, in presenting our findings at conferences, lab meetings, and in print, we have found that a list of summary questions is an effective way to communicate the items of the Wiki Quality Instruments to new audiences.

Summary questions of coding categories used in the Wiki Quality Instrument

Category *Sub-category*	**Summary question**
Information Consumption	
Course Materials	Do students come to the wiki to access academic materials?
Information Gateway	Do students come to the wiki to access links to other Web sites?
Participation	
Contribution	Does at least one student contribute, in any form, to the wiki?
Individual Pages	Does at least one student own their own page on the wiki?
Shared Pages	Does at least one pair (or group) of students own their own wiki page?
Ownership	Do student(s) serve as primary facilitator and content creator of the wiki?
Expert Thinking	
Academic Knowledge	Does at least one student complete a task requiring academic knowledge (as opposed to simply writing about hobbies or one's family)?
Information Organization	Does at least one student complete a task requiring information organization, rather than routine information retrieval?
Metacognition	Does at least on student reflect on his/her work product or process?
Crediting	Does at least one student credit their sources of his/her work?
Teacher Feedback	Do teachers provide feedback on student work?
Complex Communication	
Concatenation	Do multiple students add discrete sections of text to the same page?
Copyediting	Does at least one student copyedit text created by another student?
Co-Construction	Does at least one student substantively edit text created by another student?
Commenting	Does at least one student comment upon another student's work on the wiki?
Discussion	Do students respond to each others' comments for at least four conversational turns?
Scheduling	Do students schedule meetings or tasks?
Planning	Do students plan for future work?
New Media Literacy	
Formatting	Does at least one student use formatting elements beyond plain text?
Links	Does at least one student post a link to another page or document?
Hyperlinks	Does at least one student create links rendered as simple text or images?
Images	Does at least one student embed an image into a page?
Uploads	Does at least one student upload a document?
Multimedia	Does at least one student embed a multimedia element into a page?

Section 2: Definitions

The following key terms are referenced in the WQI decision rules:

Educator: A person acting in the capacity as a professional educator: including teachers, librarians, school administrators, IT staff, homeschool teachers, coaches, etc.

Student: Students are young people enrolled in a learning experience: classroom students, homeschool students, athletes on a team, members of a club, etc.

Others: Persons who do not fit in one of the above two categories including parents, families, and community members not employed by some kind of educational institution.

Creator: The primary content-creator or facilitator of the wiki environment. Usually the person with the most edits and the earliest edits. The person who "owns" and exerts the primarily editorial control over the content, structure, and presentation of the wiki. Each wiki should be coded as either "Student", "Educator", or "Other", but not in multiple categories.

Participants: A person or persons who are not the creator who contribute to the wiki. These are people who made some kind of direct change to the wiki (page edit, comment post, document upload, etc.), but are not the creator. Each wiki can be coded as Student, Educator, or Other, and can be coded in multiple categories.

Audience: A person or persons who are the intended viewers of a given wiki. Each wiki can be coded as Student, Educator, or Other, and can be coded in multiple categories. By default, we assume that student-created wikis have an educator audience unless we find evidence to the contrary (such as a student who creates a wiki with study guides for all of her classes intended to be used just by other students and not evaluated by an educator). We also assume that educator-created wikis with “student-facing” materials—such as naming the wiki for a course or class, posting syllabi, instructions, assignments, etc.—have a student audience. For all other audience designations, we look for specific in text or contextual references to a specific audience. For instance, to code a student-created wiki as having a student-audience, we would need evidence that the student creator intends for other students to view the wiki (“Hey guys, I hope this study guide for chapter 4 is helpful; let me know if you have suggestions!). To code an educator-created wiki as having an educator audience, we look for similar kinds of evidence (such as a page of links to lesson plans for teaching about the area of a circle).

Section 3: Decision Rules

Demographic Questions

	Indicator	Score as "0" if there is	Score as "1" if there is
Meets Criteria			
	Viewable	When the URL is entered, it returns an error because the wiki is private, deleted, or unchanged	The wiki can be viewed in a browser.
	US Based?	No evidence that wiki supports learning in U.S. based schools	Evidence that the wiki supports learning in U.S. based schools, including Department of Defense schools overseas
	K-12	No evidence that the wiki supports teaching and learning in the K-12 grades.	Evidence that the wiki supports teaching and learning in the K-12 grades.
Grade Level			
	K-5	No evidence that the wiki supports teaching and learning in the K-5 grades.	Evidence that the wiki supports teaching and learning in the K-5 grades.
	6-8	No evidence that the wiki supports teaching and learning in the 6-8 grades.	Evidence that the wiki supports teaching and learning in the 6-8 grades.
	9-12	No evidence that the wiki supports teaching and learning in the 9-12 grades.	Evidence that the wiki supports teaching and learning in the 9-12 grades.

Indicator	Score as "0" if there is	Score as "1" if there is
University	No evidence that the wiki supports teaching and learning in a post-secondary setting.	Evidence that the wiki supports teaching and learning in a post-secondary setting.

Category	Indicator	Code as “0” if…	If not “0”, code as text with the following information
Narrative	Narrative	The wiki has only been barely started with no information about the users or planned activities. (Note: sometimes even a URL is enough to begin to recognize probable users and activities)	Write a few phrases describing the usage and purpose of the wiki
Create Date	Create Date	Never. If a wiki is viewable, the create date is viewable.	Enter the date of the first version of the Front Page, which is automatically generated by PBworks.
Host	School Name	Cannot identify the hosting school or no school is involved	Write the name of the primary hosting school
	District Name	Cannot identify the hosting district or no district is involved	Write the name of the primary hosting district
	Site Name	Cannot identify another hosting institution (supra-district organization, public library,	Write the name of the primary hosting institution

	homeschool organization, etc.) or no other institution is involved	
School (2) Name	Cannot identify the secondary hosting school or no secondary school is involved	Write the name of the secondary hosting school
District (2) Name	Cannot identify the secondary hosting district or no secondary district is involved	Write the name of the secondary hosting district
Site (2) Name	Cannot identify another secondary hosting institution or no other secondary institution is involved	Write the name of the secondary hosting institution

Category	Indicator	Score as "0" if there is	Score as "1" if there is
Subject Area			
	Contained Elementary	No evidence that the wiki supports teaching and learning in a contained elementary classroom. Contained elementary refers to wikis used in elementary school classrooms where a single teacher teaches all subjects to a classroom.	Evidence that the wiki supports teaching and learning in a contained elementary classroom. Contained elementary refers to wikis used in elementary school classrooms where a single teacher teaches all subjects to a classroom. **In contained elementary wikis, it's likely that you will give a 1 for "Contained Elementary" as well as a 1 in several subject areas.**
	English/ Language Arts	No evidence that the wiki supports teaching and learning in English or Language Arts subjects. ELA classes are devoted to the study of English, reading, writing, literature, poetry, prose, textual analysis of drama, spelling, and so forth.	Evidence that the wiki supports teaching and learning in English or Language Arts subjects. ELA classes are devoted to the study of English, reading, writing, literature, poetry, prose, textual analysis of drama, spelling, and so forth.
	Math	No evidence that the wiki supports teaching and learning in Mathematics.	Evidence that the wiki supports teaching and learning in Mathematics.

Science and Engineering	No evidence that the wiki supports teaching and learning in the natural and physical sciences- biology, chemistry, physics, geology, etc—or engineering courses.	Evidence that the wiki supports teaching and learning in the natural and physical sciences- biology, chemistry, physics, geology, etc—or engineering courses..
English as a Second Language	No evidence that the wiki supports teaching and learning of English to non-native English speakers. (Sometimes also called English as a Foreign Language or English Language Learners).	Evidence that the wiki supports teaching and learning of English to non-native English speakers. (Sometimes also called English as a Foreign Language or English Language Learners).
Social Studies	No evidence that the wiki supports teaching and learning in the Social Studies, including history, religion, psychology, economics, geography, and other social sciences.	Evidence that the wiki supports teaching and learning in the Social Studies, including history, religion, psychology, economics, geography, and other social sciences.
Computer Science and Technology	No evidence that the wiki supports teaching and learning in Computer, Computer Science, Programming (Java, C++, etc), Technology Applications, or Typing Classes, or teaching and learning about educational Web/technology applications independent of an academic subject area.	Evidence that the wiki supports teaching and learning in Computer, Computer Science, Programming (Java, C++, etc), Technology Applications, or Typing Classes. Also includes wikis that support teaching and learning about educational Web/technology applications independent of an academic subject area.
Modern Foreign Language	No evidence that the wiki supports teaching and learning in Modern Foreign Languages such as Spanish, French, German, Chinese, etc.	Evidence that the wiki supports teaching and learning in Modern Foreign Languages such as Spanish, French, German, Chinese, etc.

Classic Languages	No evidence that the wiki supports teaching and learning in Classical Languages such as Latin and Greek	Evidence that the wiki supports teaching and learning in Classical Languages such as Latin and Greek
Art	No evidence that the wiki supports teaching and learning in any form of visual, studio, or performing arts.	Evidence that the wiki supports teaching and learning in any form of visual, studio, or performing arts.
Business	No evidence that the wiki supports teaching and learning in business classes.	Evidence that the wiki supports teaching and learning in business classes.
Health/Physical Education	No evidence that the wiki supports teaching and learning in health classes, sexual education, physical education, or extra-curricular sports teams.	Evidence that the wiki supports teaching and learning in health classes, sexual education, physical education, or extra-curricular sports teams.
Education	No evidence that the wiki supports teaching and learning in education classes for teachers or pre-service teachers.	Evidence that the wiki supports teaching and learning in education classes for teachers or pre-service teachers.
Library	No evidence that the wiki supports a library site or teaching and learning in library or media studies courses.	Evidence that the wiki supports a library site or teaching and learning in library or media studies courses.
Other	No evidence that the wiki supports teaching and learning in a subject area not defined above.	Evidence that the wiki supports teaching and learning in a subject area not defined above.

Creator			
	Student	No evidence that a student is the primary content-creator or facilitator of the wiki	Evidence that a student is the primary content-creator of facilitator of the wiki. **Common evidence in this category includes (but is not limited too) the creator identifying themselves as a student; academic work which is in response to a particular prompt; naming or identifying the wiki as in service of a student or a working group; grammar, spelling and syntax characteristic of young people.**
	Educator	No evidence that an educator is the primary content-creator or facilitator of the wiki.	Evidence that an educator is the primary content-creator or facilitator of the wiki. **Common evidence includes: self-identifying as a teacher, posting instructional materials, posting instructions for academic work, naming or identifying a wiki as in service of a course or class rather than for a student or working group.**
	Other	No evidence that an educator is the primary content-creator or facilitator of the wiki.	Evidence that a non-student, non-educator is the primary content-creator or facilitator of the wiki. **Rare**.

Participant			
	Student	No evidence that any student (not counting the creator) contributes to the wiki by contributing to a page edit, leaving a comment, or uploading a document.	Evidence that at least one student (not counting the creator) contributes to the wiki by contributing to a page edit, leaving a comment, or uploading a document.
	Educator	No evidence that any educator (not counting the creator) contributes to the wiki by contributing to a page edit, leaving a comment, or uploading a document.	Evidence that at least one educator (not counting the creator) contributes to the wiki by contributing to a page edit, leaving a comment, or uploading a document.
	Other	No evidence that any non-student, non-educator (not counting the creator) contributes to the wiki by contributing to a page edit, leaving a comment, or uploading a document.	Evidence that at least one non-student, non-educator (not counting the creator) contributes to the wiki by making a page edit, leaving a comment, or uploading a document.

Audience			
	Student	No evidence that students are meant to be the audience of the wiki.	Evidence that students are meant to be the audience of the wiki**. Most educator-created wikis have students as their audience. By default, if an educator is posting academic materials or instructions for academic work, we assume the audience of the wiki is students. If a student is the wiki creator, only code 1 here is there is evidence that the wiki is intended for viewing by other students. Evidence would include a specific reference to a student audience.**
	Educator	No evidence that educators are meant to be the audience of the wiki.	Evidence that educators are meant to be the audience of the wiki. **Unless there is clear evidence to the contrary, we assume that all sites created by students have educators as an audience. If an educator is creating a resource site for other educators, code this category as a 1.**
	Other	No evidence that others are meant to be the audience of the wiki.	Evidence that others are meant to be the audience of the wiki. **Only code 1 if there is explicit evidence that parents, experts or others are intended to be an audience for the site.**

Quality Questions

Category	Indicator	Score as "0" if there is	Score as "1" if there is
Information Gateway			
	Course Materials	No evidence that instructional or content materials are posted on the wiki.	Evidence that instructional or content materials are posted on the wiki by teachers for viewing by students or posted by students for viewing by other students (i.e. do not code 1 if a student posts instructional or content materials for teacher viewing, other viewing, or for the student's own benefit)
	Information Gateway	No evidence that links to resources that are beyond the wiki subdomain are posted on the wiki by teachers for viewing by students or posted by students for viewing by other students.	Evidence that links are posted on the wiki to resources that are beyond the wiki subdomain by teachers for viewing by students or posted by students for viewing by other students (i.e. do not code 1 if a student posts links to external sites for teacher viewing, other viewing, or for the student's own benefit)

Category	Indicator	Score as "0" if there is	Score as "1" if there is
Participation			
	Contribution	No evidence that students contribute by making edits or changes (including but not limited to adding text, uploading files, etc.) to the wiki	Evidence that students contribute by making edits or changes (including but not limited to adding text, posting comments, uploading files, etc.) to the wiki
	Ownership	No evidence that a student (or students) owns the wiki. Wiki “owners” are the wiki’s creator, primary facilitator or primary content producer.	Evidence that a student (or students) owns the wiki. Wiki “owners” are the creator, primary facilitator or primary content producer for the wiki.
	Individual Pages	No evidence that at least one student owns an individual page. Wiki page “owners” are the creator, primary facilitator or content producer.	Evidence that at least one student owns an individual page. Wiki page “owners” are the creator, primary facilitator or content producer for the page.
	Shared Pages	No evidence that at least two students co-own a single shared page. Wiki page “owners” are the creator, primary facilitator or content producer for the page.	Evidence that at least two students co-own a single shared page. Wiki page “owners” are the creator, primary facilitator or content producer for the page.

Category	Indicator	Score as "0" if there is	Score as "1" if there is
Expert Thinking		**EVIDENCE IN EXPERT THINKING CATEGORIES CAN COME FROM UPLOADS**	
	Content Required	No evidence that students complete tasks requiring academic content knowledge, or information and skills that are taught as part of the school curriculum.	Evidence that at least one student attempts to complete a task on the wiki requiring academic content knowledge, or information and skills that are taught as part of the school curriculum. Simple reading and writing do not count here, although specialized forms of reading (such as grammatical parsing) or writing (such as writing poetry or drama) do count.
	Knowledge organization	No evidence that students complete tasks requiring analyzing, organizing, or synthesizing information.	Evidence that at least one student attempts to complete tasks on the wiki requiring analyzing, organizing, or synthesizing information.
	Self-Reflection	No evidence that students use the wiki to reflect on their thinking process or work products.	Evidence that at least one student uses the wiki to reflect on his/her thinking process or work product. The reflective exercise must occur on the wiki, but the object of reflection does not have to be on the wiki.
	Crediting	No evidence that students credit the supports of their work products.	Evidence that at least one student attempts to credit the supports of his/her work product on the wiki.
	Teacher Feedback	No evidence that there is **teacher** feedback on the wiki to comment on student products.	Evidence that that there is ***teacher*** feedback on the wiki that comments on the thinking process or work products of students.

Category Indicator	Score as "0" if there is	Score as "1" if there is
Complex Communication		
Commenting	No evidence that students comment upon the work of another wiki user (Student or non-student).	Evidence that at least one student comments upon the work of another wiki user (Student or non-student).
Discussion	No evidence that students sustain a conversation of at least four turns with other users (student or non-student)	Evidence that at least one student engages in a discussion with at least four conversational turns with another user
Concatenation	No evidence that more than one student contributes discrete text or content to a page.	Evidence that at least two students contribute discrete text or content to a page. Evidence of concatenation can come from bylines attributed to discrete text or from edit histories.
Copyediting	No evidence that one student corrects the grammar, spelling, punctuation, or syntax of another user’s discretely owned text or content.	Evidence that at least one student corrects the grammar, spelling, punctuation, or syntax of another user’s discretely owned text or content. Evidence generally must come from edit histories with clear identities.

Co-Construction	No evidence that a student substantively edits the work of another user working on the same page such that the text is no longer discretely owned.	At least one student substantively edits the work of another user working on the same page such that the text is no longer "discretely" owned. Evidence generally must come from edit histories with clear identities.
Scheduling	No evidence that a student participates in a scheduling activity on the wiki, such as posting their name to a list of times or responsibilities.	Evidence that a student participates in a scheduling activity on the wiki, such as posting their name to a list of times or responsibilities.
Student planning	No evidence that students discuss and develop strategies on the wiki for completing a work product (on or off wiki) with other students.	Evidence that at least one student discusses and develops strategies on the wiki for completing a work product (on or off wiki) with at least one other student.

Category	Indicator	Score as "0" if there is	Score as "1" if there is
New Media Literacy			
	Formatting	No evidence that students use formatting such as colored text, boldface, italics, and so forth on a wiki page. (Formatting on attached documents does not count).	Evidence that at least one student uses formatting such as colored text, boldface, italics, on a wiki page. (Formatting on attached documents does not count).
	Links	No evidence that students add Web link to a wiki page.	Evidence that at least one student adds Web links to a wiki page. The links can be to other wiki pages, external Web sites, or to uploaded documents. (Links within attached documents do not count). For example: http://www.google.com.
	Hyperlinking	No evidence that students add links to non-URL text or images on a wiki page.	Evidence that at least one student adds a link to non-URL text or images on a wiki page. For example: Google
	Images	No evidence that students embed images to a wiki page.	Evidence that at least one student embeds an image into a wiki page. You must be able to see the image on a page. Merely uploading a file should be coded under “Upload”. Simple links should be coded under “Links.”

Multimedia	No evidence that students embed sounds, videos, or other multimedia files to a wiki page.	Evidence that at least one student embeds sounds, videos, or other multimedia files to a wiki page. (Multimedia files within attached documents do not count. The image or multimedia application must appear on the page as an embed frame. Merely uploading a file should be coded under "Upload". Simple links should be coded under "Links.").
Upload	No evidence that students upload files to the wiki.	Evidence that at least one student uploads a file to the wiki.

Final Overall Rating Questions

For the final categories, answer the following questions on a 1-7 scale (7 being highest)

Overall Participation: Overall, to what extent do students participate in this wiki?

No participation 1	Minimal participation 2	A little participation 3	Some participation 4	A fair amount of participation 5	Considerable participation 6	Extensive participation 7

Overall Expert Thinking: Overall, to what extent does this wiki promote the development of students' expert thinking skills—such as ill-structured problem solving, critical thinking, creative problem solving, metacognition, initiative taking?

No expert thinking 1	Minimal expert thinking 2	A little expert thinking 3	Some expert thinking 4	A fair amount of expert thinking 5	Considerable expert thinking 6	Extensive expert thinking 7

Overall Complex Communication : Overall, to what extent does this wiki promote the development of students' complex communication skills—such as communication and collaboration?

No complex communication 1	Minimal complex communication 2	A little complex communication 3	Some complex communication 4	A fair amount of complex communication 5	Considerable complex communication 6	Extensive complex communication 7

Overall New Media Literacy: Overall, to what extent does this wiki promote the development of students' new media literacy skills-the skills necessary to critically consume and produce multimedia content?

No new media literacy 1	Minimal new media literacy 2	A little new media literacy 3	Some new media literacy 4	A fair amount of new media literacy 5	Considerable new media literacy 6	Extensive new media literacy 7

Part II: Content Analysis and Training Guidelines and Protocols

Section 1: Overview of Content Analysis Protocols
Section 2: Wiki Sampling
Section 3: Content Analysis Work Flow
Section 4: Evaluating Wiki Edit Histories
Section 5: Ethical Concerns in Online Content Analysis and IRB Considerations
Section 6: Overall Ratings
Section 7: Training Protocols
Section 8: Conclusion

Section 1: Overview of Content Analysis Protocols

The wikis that we study are extremely diverse. They are used with elementary schools through high schools, in nearly every subject area imaginable, and for a wide variety of educational purposes. They range in size and complexity from a single page with no revisions to wikis with hundreds of pages revised thousands of times. Characterizing the activity on wikis accurately is very challenging work. In this section, we present the current strategies that we are using to meet these challenges. Readers of some of our published articles may find some discrepancies between the procedures described here and our published procedures. The protocols listed here reflect our most recent, most refined thinking.

In our first round of coding, researchers evaluate the demographic features of a wiki, which we treat as time-invariant. Two coders evaluate each wiki. First, they determine the wiki's eligibility for our studies by confirming that the wiki is visible (not private, deleted, or unchanged[1]), used in the United States, and used in K-12 settings. The same coders then determine the subject area(s), grade level(s), and hosting school(s), district(s) or other site(s) of the wiki. A third coder reconciles disagreements. We then provide this information to two additional coders who evaluate wiki quality. (We attempt to have the same people who evaluated wiki demography evaluate wiki quality, since they can often do the quality coding more quickly than someone who needs to examine a wiki for the first time. Sometimes the timing does not work out—some people work faster or have more time than others—so this is a preference rather than a rule.)

To evaluate wiki quality, we use the Wiki Quality Instrument (WQI). The WQI contains 24 dichotomous items that probe for the presence or absence of behaviors on the wiki that provide opportunities for students to develop 21st century skills. These 24 items are in five subdomains: Information Consumption (2 items), Participation (4 items), Expert Thinking (5 items), Complex Communication (7 items), and New Media Literacy (6 items). We measure wiki quality at days 1, 7, 14, 30, 60, 100, and 400. Two coders evaluate wiki quality for each wiki at each of these time points. A third coder reconciles disagreements.

To determine a composite wiki-quality score, we sum the values of the 24 WQI items. To determine subdomain wiki-quality scores, we sum the values of the items in each of the five subdomains. These scores are then used as outcome measures in our analyses.

Section 2: Wiki Sampling

[1] Some wikis are created and then never viewed at all by the creator, and when a coder visits the URL of one of these wikis they receive an error message. Some wikis are created and then viewed by the creator, and our raters could view these, even though they were unchanged.

We have gathered three wiki samples from PBworks.com, each using different strategies, in order to conduct different studies to answer different kinds of research questions.

In our first sample, we sought to evaluate a representative sample of all wikis created from the founding of PBworks in June of 2005 through August of 2008. PBworks provided us with a list of the URLs for all the publicly-viewable, education-related wikis. We assigned each of these 179,851 wikis an ID number, and we used random.org to draw a non-repeating series of 1,799 numbers from the larger set. We separated out 411 U.S. K-12 wikis from this random sample, and we then analyzed this subset in our first research study.

In our second sample, we sought the ability to identify a representative sample of wikis very close to their creation date so that we could survey wiki creators soon after they started their wiki. To do this, in September of 2010, we had PBworks send us a list of all recently-created, publicly-viewable, education-related wiki URLs every two weeks. These wiki creators received an automated survey solicitation, generated by PBworks, as soon as they signed up for a wiki. We then sampled 500 wikis at random, every two weeks, for eight weeks (giving us a sample of 2,000 wikis out of approximately 20,000 publicly-viewable, education-related wikis created during this period). As soon as we received the list, we then sent each wiki a follow-up solicitation to participate in our survey. Our response rate for this sample was fairly low, we received only 80 survey responses that included a URL allowing us to link teacher attitudes with wikis. Thus, we collected a larger third sample.

In our third sample, we gathered a convenience sample of all wiki-creators who had completed our survey. All PBworks educational wiki creators from June 2010 through December of 2010 received an automated survey solicitation which included the option to provide us with their URL, so we could include them in our study. Through our surveys we received 510 URLs. Since we do not know precisely how many U.S., K-12 wikis were created during this period, it is impossible to precisely know the response rate, but it is low, probably below 10%. If the shared-URLs were publicly viewable, they were included in our survey. If the URLs we received were for private wikis, we sent a solicitation inviting the host to join our study by adding us to their wiki community.

Section 3: Content Analysis Work Flow

Coding thousands of wikis at multiple time points with multiple raters presents a set of complicated logistical challenges. Through experimenting with a variety of protocols and processes, we have settled on a set of strategies to manage these challenges. In this section we describe how we organize our wikis and conduct our demographic and quality analyses.

Assigning IDs

First, when we receive the “population lists” of wikis from PBworks, we assign every wiki a unique ID. This identification number is essential; it allows us to have multiple researchers simultaneously analyze different aspects of a wiki, and then bring these analyses together in a common dataset. After assigning ID's to wikis and drawing samples from the population, we organize lists of sampled wikis into spreadsheets for further analysis.

Organizing Coding Activities with Spreadsheets

We organize our coding activities around a large set of short Excel worksheets. Each sheet is designed so that it includes only those wikis and wiki time-periods that a coder is expected to work on, only the information that she needs for the task, and only those blank columns which she needs to complete. (We experimented with having large sheets and assigning coders to only work on parts of them. For instance, we created worksheets with 500 rows and would assign a coder to complete rows 101-200. This led to errors, with coders completing the wrong rows, so we settled on a system where coders only received one sheet at a time that included exactly the amount of work they were supposed to complete). Most of our coders are Masters students who work 8-10 hours per week, so sheets are designed to take 1-2 weeks to complete. In the sections below, we describe how we create and organize these various spreadsheets to conduct our coding exercises. Examples are available from the first author.

Wiki Coding Round 1: Demographic Coding

In the first round of demographic coding, we divide our wikis into sets of 100. We create a series of spreadsheets where each sheet contains 100 rows, each corresponding to one wiki. Each sheet has several pre-populated columns of information, such as the URL, the creation date supplied by PBworks, and the ID. Each sheet also has one column header for each of the demographic questions in the WQI. While some of these items require text answers (such as a short narrative of the wiki's purpose or the name of the wiki's hosting school), most are dichotomous items (does the wiki serve K-5 students? 6-8? 9-12?). Therefore, we can easily validate the data of these dichotomous items by ensuring that all fields include only 1s or 0s, with no blanks or other characters. The list of items and decision rules for each item can be found in the Wiki Quality Instrument. Each sheet is given to two different coders for the initial round of demographic analysis.[2]

We treat the demographic information for each wiki as time-invariant. Thus, we do not track the exact time in which students begin participating in a wiki created by an educator, we simply record the wiki as having student participants. This means that for our demographic coding sheets, each row corresponds to a single wiki.

Identifying demographic information on a wiki often involves a certain amount of "detective" work. In some wikis, all of the demographic information is very clearly presented. For instance, a teacher might create a wiki for her students and indicate the course the wiki is used in, the school she works at, the grade level of her students, and so forth. In these circumstances, identifying the demographic characteristics of the wiki is quite simple. In many cases, however, this analysis is much less straightforward.

For instance, we encounter a large proportion of wikis where the default wiki text is unchanged or changed in only trivial ways. We still hope to understand as much as we can about these wikis, so we can compare more and less successful wiki learning environments. Therefore, we invest quite of bit of time in examining whatever snippets of content are available to discover as much as we can about each wiki. For instance,

[2] Sample sheets are available by request from the authors. We have not posted them here since we have decided not to repost URLs of wikis from our study.

teachers often use their real names as their PBworks user IDs, so sometimes we can search for these names online and identify the school where a teacher works and the classes he or she teachers. This works better with more uncommon names. Often, teachers and students put their email address on a wiki, which can help us identify where they are from and what subjects the wiki is meant to support. Sometimes teachers will have links to other online learning environments, like their blogs or class websites, that allow coders to identify users and schools. Teachers sometimes use their school initials in the wiki URL, so if we find the initials CRLS in a URL, we can search for that combination of letters and find that there is a school called Cambridge Rindge and Latin in Cambridge, Massachusetts. Whenever possible, we use multiple sources of information like this to triangulate and corroborate our findings.

For experienced research assistants, conducting the demographic coding on a spreadsheet containing 100 wikis typically takes between 5 and 12 hours. In our budgeting and time planning, we assume that demographic coding the average sheet of 100 wikis would take about 8 hours on average, or approximately 5 minutes per wiki. Many wikis can be coded very quickly, in a minute or two. Wikis created in foreign countries or in higher education settings are not eligible for our study; these can be coded very quickly. Some wikis contain content which clearly labels the activity on the wiki or contain almost no content at all making identification impossible; these are also simple to code. These are balanced by wikis with many pages or with limited information that can require as many as 30 minutes to examine and explore.

Wiki Coding Round Two: Demographic Coding Reconcile

Since two coders complete each demographic coding task, we have a third rater reconcile disagreements. To reconcile the two original worksheets, we take each completed demographic sheet and copy the data into a new Excel workbook. One sheet is named CODER1, and the second sheet is named CODER2. We then create a third sheet which we use to reconcile disagreements.

For all of the columns that record the values of dichotomous items (grade level, subject, users), we use a series of IF functions in columns to identify disagreements. For instance, assume that column B records whether or not a wiki is visible, with cells coded "0" for not visible (set to private or deleted) or "1" for visible. For cell B2 on the reconcile worksheet, we use the formula IF(CODER1!B2=CODER2!B2, CODER1!B2, "X"). This formula reconciles the B2 columns from both coder sheets such that if two coders agree on a rating, the rating stands. If they disagree, it is marked with an X.

For all of the columns that record text values, such as the wiki narratives and the names of hosting institutions, we simply transfer this information onto the third reconcile sheet. The narratives are not reconciled. They are there for the reference of future coders. The school, district, and/or other hosting site information is also transferred to the third reconcile sheet, and an additional set of columns are generated to be completed by the third rater to create a final reconciled list of hosting institutions.

This third reconcile sheet is generated entirely using Excel formulas, so to create a copy to send to the third rater, we copy the entirety of the third sheet (comprised of formulas) onto the fourth sheet using the "Paste Values" options, so the fourth sheet is comprised entirely of values. We then use conditional formatting to mark all Xs with a red background. This is then sent to the third rater, who is tasked with turning all of the red Xs into 1s and 0s and with determining the final list of hosting institutions.

The time required for usage reconciling varies considerably, both with the complexity of the wikis in any particular sheet as well as the degree of difference between two raters. Typically, in a set of 100 wikis, over 50 will have perfect agreement between the two raters (mostly on wikis that do not meet criteria for the study), so the third rater is only examining half of the wikis. Another substantial portion of wikis, typically about 33, will only have a small number of disagreements. The remaining 17 wikis will have more substantial disagreements. Experienced coders typically took approximately 3 hours to reconcile a sheet. Thus, if the initial two raters take approximately 8 hours each to code the demographic questions from the WQI, and a third rater takes approximately 3 hours to reconcile, then each set of 100 wikis will take approximately 20 worker-hours for the demographic coding. Thus, a sample of 2000 wikis might take about 400 worker hours. In our group, with 10 researchers working about 8 hours a week at $16/hour, that would take approximately five weeks to complete and cost about $6400.

Once reconciled, this sheet will consist of 100 rows, each representing a wiki, where each row records demographic information about the wiki. We then use the filtering option to remove all wikis that are not eligible for our study (private, deleted, not U.S., not K-12) from the set. Typically, approximately 60% of wiki URLs out of a random sample of publicly-viewable, education-related wikis will prove to be ineligible for our study. This filtered set of eligible is saved to be prepared for quality coding.

Wiki Coding Round 3: Quality Coding

We believe that quality is a time-varying feature of wiki learning environments, and therefore we evaluate wikis at multiple time points. In our most recent studies, we evaluate wikis at 1, 7, 14, 30, 60, 100, and 400 days. We provide a more complete description of our reasons for choosing these time points in Part IV: Developing the WQI Protocols, but we summarize our rationale here. We know from prior research that the median lifetime of wikis in U.S. public schools is about 13 days, that on average most wiki activity happens within the first weeks after a wiki's creation, and that on average wiki-quality growth is greatest within the first week. Since capturing wiki-quality growth in the first two weeks of a wiki's lifetime is vital, we take measurements with the WQI at days 1, 7, and 14. In our most recent studies, we also have taken measurements at days 30 and 60, which represent two and four times the median lifetime, and at day 100, which is approximately the duration of one semester. If we needed to reduce the number of measurements in order to save costs, we would probably eliminate the measurement at day 60 first and then day at 30. We also take measurements at day 400, since we know that many wikis experience a surge of activity after approximately one year.

To generate our quality coding sheets, we take our completed demography sheets and we use a SAS routine to generate one row for each of the occasions of measurement for each wiki. The SAS routine takes the wiki's creation date, and adds 1 day for the day one measure, adds 7 days for the day seven measure, and so forth. Thus if a coder is assigned to evaluate a wiki created on Feb. 1 on its seventh day, they will know to evaluate all edits made through Feb. 8. Since most reconciled demography sheets contain approximately 40 eligible wikis, if we conduct an analysis that includes seven occasions of measurement, our quality sheets will typically have approximately 280 rows.

To code the quality sheets, coders first examine each wiki and determine the last day it was changed.[3] The wiki is then coded for all occasions of measurement that capture changes in the wiki up to day 400. For instance, if a wiki's last change is on its 16th day, it will be coded for days 1, 7, 14, and 30. If a wiki's last change is on its 99th day, it will be coded on days 1, 7, 14, 30, 60, and 100. If its last change is on its 645th day, it will be coded on days 1, 7, 14, 30, 60, 100, and 400. The quality coding sheets that we generate have a row for every occasion of measurement, and coders manually determine which rows are to be completed and which are to be left blank. On average, we code approximately 4 time periods per wiki. Thus, on a sheet with 40 wikis, we are likely to code approximate 160 rows.

At each occasion of measurement we identify behaviors on the wiki that provide opportunities for students to develop 21st century skills that have occurred up to the occasion of measurement. Therefore, when we evaluate wiki quality at day 14, we evaluate all opportunities for students to develop 21st century skills up through day 14. This means that any of our quality items coded "1" at day 1 will also be coded "1" at all subsequent days. In our coding scheme, opportunities for 21st century skill development cannot disappear. This means that wiki-quality scores are monotonic.

---

[3] We have experimented with developing computational tools for determining a wiki's creation date. We have found that a small number of districts and schools have institutional wiki creation processes. In these cases, API calls to the PBworks data warehouse for the wiki creation date can return dates for when a group of wiki subdomains are named and reserved, rather than when the wiki is actually first generated. Thus we manually check each wiki creation date.

Coders use the WQI to identify the presence or absence of 24 types of behaviors that commonly occur on wikis that provide opportunities for students to develop 21st century skills. The behaviors are detailed in the decision rules of the quality items for the WQI.

In order to be able to evaluate how wiki usage changes over time, our coders must be able to evaluate every revision to every page and file on the wiki. In the next section, we detail the strategies that our coders use to conduct this historical analysis.

In addition to coding the 24 quality items of the WQI, our research assistants subjectively rate wikis based on the degree to which they feel that the wiki provides opportunities for participation, expert thinking, complex communication, and new media literacy development. These subjective ratings are never reconciled. They are described further in Section 6: Overall Ratings.

In terms of timing, most coders can code approximately 15 occasions of measurement in one hour, on average (again, some wikis take only a few seconds, and others can take hours). Out of 100 randomly sampled wikis, 40 will be eligible for our study. Each wiki will require on average 4 time periods to be coded, meaning that we typically need to code 160 rows. If coders complete approximately 15 time periods in one hour, then the typical sheet takes between 10 and 12 hours for one coder to evaluate the quality items from the WQI.

Wiki Coding Round 4: Quality Reconciling

After two coders have conducted the quality coding for each wiki, their two sheets are reconciled using the IF function in Excel as described above. All disagreements are marked with an X and highlighted in red using conditional formatting. A third rater then reconciles these quality codes, turning all X's into 1's and 0's. Again, about half of all rows will be in perfect agreement, and then other half will require some form of reconciling. We found that researchers typically took approximately 8 hours to reconcile a typical sheet.

Thus with two people coding, each requiring 12 hours, and approximately 8 hours of reconciling, a typical sheet of 40 wikis (separated out from the original 100 wikis) coded at seven time periods requires approximately 32 worker-hours to complete. Therefore, to code 800 wikis (separated out from the original 2000) requires approximately 640 worker-hours. In our group, with 10 researchers working about 10 hours a week, that would take approximately eight weeks to complete and at $16/hour, cost about $10200. The entire process, therefore, takes a team of 10 trained researchers approximately three months to evaluate 2000 wikis and costs approximately $17,000 at $16/hour.

Merging Coding Data and Other Data

Once all quality coding is completed, all of the reconciled quality sheets can be merged into a dataset which includes all of the demographic and quality codes for each occasion of measurement for each eligible wiki.

This dataset then is merged with two other sources of data. In order to obtain demographic data about hosting schools, we use data from the National Center for Education Statistics Common Core of Data. To obtain this data, first, a researcher examines our records of the hosting institution for each wiki. These hosting institutions can be public schools, independent schools, districts, town libraries, district consortia (like the BOCES in New York or the Area Education Associations in Iowa), or other institutions. For wikis hosted by public schools or districts, we obtain their NCES School ID and NCES District ID from the website http://nces.ed.gov/ccd/schoolsearch/index.asp. We then use these ID numbers to obtain data from three data files hosted at http://nces.ed.gov/ccd/ccddata.asp: the Public Elementary/Secondary Universe Survey Data (which contains data about individual schools) and the Local Education Agency (School District) Universe Survey and Local Education Agency (School District) Finance Survey (which contain data about school districts). Using these ID numbers, we merge the demographic data about schools and districts—such as the percentage of students in a wiki's hosting school eligible for Free and Reduced Priced Lunches—into our wiki dataset.

For some of our samples, we also have teacher surveys. For a time, PBworks agreed to automatically send a survey solicitation to all creators of education-related wikis. We are dependent upon survey takers entering the URL of their wiki into a question in our survey to identify the wiki associated with each survey takers. We use that URL as the link between our wiki dataset and the survey data.

Outline summary of workflow

1. Obtain population level data from PBworks
    a. Assign IDs
2. Take random sample of wikis
    a. Use random.org to draw a non-repeating series of IDs from the population list
3. Code for WQI demographic items
    a. Two coders evaluate demographic items
    b. Identify disagreements using Excel's IF function
    c. Third coder reconciles disagreements
    d. Filter out ineligible wikis
4. Code for WQI quality items
    a. Two coders evaluate quality items
        i. Two coders subjectively rate wiki quality, and these ratings are never reconciled
    b. Identify quality coding disagreements using Excel's IF function
    c. Third coder reconciles disagreements
5. Obtain school and district level demographic data
    a. Coder uses school and district names to identify NCES School and District ID numbers
    b. Using NCES ID numbers, merge wiki-quality data with school and district demographic data
6. Obtain teacher survey measures
    a. Use wiki URL's provided by teacher survey takers to link wiki-quality data with data on teacher attitudes and practices

Section 4: Evaluating Wiki-Edit Histories

One of the signature features of our research approach is that we leverage the fact that wikis preserve a real time history of every revision to every page. For instance, if we view only the most recent version of a wiki page, it is impossible to precisely determine the kinds of collaborative behaviors that may be responsible for creating that page. However, if we evaluate every revision for a wiki page, we can determine how multiple contributors work together in varying ways to co-construct the content of the page. As a research team, we have developed several strategies for evaluating wikis and wiki edit histories.

When evaluating a wiki for the first time, most coders will first make a holistic evaluation of the wiki. They will use the navigation settings on the right-hand sidebar of the wiki to review the current versions of each page, to see the kinds of files that are uploaded, and to evaluate the navigational structure of the page. They then use a number of different approaches to begin to probe the historical record of each wiki.

Each wiki includes a *Recent Activity* link, which takes viewers to an automatically populated page that shows all of the recent activity on the wiki: new pages created, new page revisions, new comments, and new files uploaded. If wikis have very few changes, this link may provide a comprehensive list of the total history of the wiki. For wikis with many pages and changes, this link might only provide a tiny fraction of the total wiki activity. Nonetheless, it always provides a useful overview of the patterns of activity in the recent history of the wiki.[4]

Within the PBworks navigation system, viewers can also follow the *Pages and Files* link to a listing of all pages created on the wiki and all files uploaded. For each page and file, there is a link to a list of the complete revision history of each page. In the current Graphical User Interface (GUI), this link can be found by mousing over a page or file until the *More* button appears, clicking the more button, and the clicking the link which says *x revisions*, where "*x*" is the number of revisions to the page. Following this link, the viewer will encounter a list of page (or file) revisions continuously recorded to the second.

The most efficient way to manually browse all of the revisions for a page is to use the following procedure: First, a researcher scrolls down to the first version of the page. She then right-clicks the link to this version and then chooses to *Open in a New Tab*. She will then have a new browser tab with the first version of the page. (It is important for coders to realize that when looking at historical page revisions, the comments are not included in this view. Comments are only viewable when looking at the present version of a page through the regular browser interface.) Next, she right-clicks on the link to the second version of the page, and she opens it in a new tab. She repeats this process for every revision. Her browser will now have a set of tabs where each tab renders a version of the wiki page in chronological order. By clicking on each tab sequentially, the coder can evaluate the changes in the wiki history.

[4] The *Recent Activity* link shows links by month and date and not by year, which can cause confusion when wikis have not been edited for several years. A review of the page histories, described in the following paragraphs, can resolve this potential confusion.

This protocol needs to be modified if coders are evaluating a wiki *up to a certain occasion of measurement*. For instance, if a coder is only evaluating a wiki through its first 30 days of changes, then the coder first needs to calculate the correct date to stop evaluating changes. In our workflow process, the dates of occasions of measurement are produced along with the coding sheets for wiki-quality coding. Thus if a wiki is created on Feb 14th, the coder knows that the day 1 measurement should evaluate all changes through Feb. 15, the day 7 measurement through Feb. 21, the day 14 measurement through Feb. 28, and so forth. With this information, the coder should only open new tabs for revisions that occur before the appropriate cutoff date.

This process can be quite cumbersome, so we developed our own browser interface to evaluate wiki edit histories: the Wiki-Coding Tool (WCT) (http://tool.edtechresearcher.com/code/). The problem with the PBworks Pages and Files interface is that *time* is nested within *pages.* That is, the edit histories for each page are recorded under the name of each page. Comparing wikis at varying time points using this model is quite difficult. Thus, we created the WCT to organize wikis such that *pages* are nested within *time*. That is, the WCT allows a coder to select an occasion of measurement, such as day 7, and to restrict her view to only those pages and revisions that occurred up to or before the wiki's seventh day.

The WCT uses the PBworks Applied Programming Interface (API) in order to "rearrange" the page revisions for each wiki. APIs on websites are the programming languages that computers use to query websites, in contrast to the navigation buttons and links that humans use in their browsers to query websites. To begin using the WCT, coders enter the URL for a PBworks wiki. Coders then choose an occasion of measurement for the wiki, and they are presented with a dropdown list of all pages that had been created up to that occasion of measurement. When a coder selects a page, she is then presented with the most recent version of the page. There are also “forward” and “back” buttons that allow her to quickly scroll between page revisions from the original version through the final revision before the cut-off time of the occasion of measurement. This system is considerably easier than the process of manually opening every revision to every page.

The WCT illustrates a crucial point for the future of education research in online learning environments: online learning platforms are designed to facilitate entering content. The kinds of navigational structures that facilitate content creation are not necessarily well-suited for evaluation of content creation processes. In our case, PBworks makes it easy for multiple people to contribute to wiki pages, but it is difficult to examine those contributions over time. The development of the WCT was our attempt to resolve this problem, and it suggests to us that education researchers will need to develop expertise in the years ahead in using APIs to reorganize online learning environment data with an eye towards content analysis rather than content contribution.

The WCT has a number of limitations. Periodically, PBworks has changed its URL structure or other features of its service, and these changes have rendered the WCT unusable for periods of time as we re-program the WCT. Also, since the WCT calls up historical versions of pages, coders still need to evaluate the entire wiki to find comments, which are not rendered on historical page revisions. We are still refining the

WCT so that it can evaluate all PBworks wikis; currently it cannot evaluate private wikis, even if we have been invited, and it cannot evaluate wikis with unusual characters in its page names. These are tractable problems, but they remain unresolved.

In summary, our coders used a variety of strategies to evaluate wikis and their edit histories. These strategies included browsing the navigation of each wiki, examining the recent activity pages generated automatically by PBworks, and using the PBworks Pages and Files interface and our own Wiki Coding Tool to conduct a detailed evaluation of every page of every wiki.

Section 5: Ethical Concerns in Online Content Analysis and IRB Considerations

The widespread availability of new forms of online data has created a new set of ethical challenges for educational researchers and Institutional Review Boards (IRB). What kinds of protections are necessary to ensure that participants in online learning environments can be kept safe from harm as we research those learning environments? How can researchers educate IRB staff about these new environments so that research can progress and new methods can be used while protecting research subjects? These are open questions as new methods tackle new technologies.

For most of our research methods, our IRB and subject consent protocols were quite typical to other forms of educational research. When we conducted interviews with faculty, we sought consent from schools and teachers. When we conducted interviews with students, we sought consent from schools, parents, and students. Our surveys solicited teacher consent. For our in-class observations, we followed our IRB protocol of requiring teacher and principal consent. Some districts required district-level consent from the central office, and some schools required parental consent for each student in a classroom.

For our analysis of publicly-viewable wikis, we were not required to gather consent. Our IRB determined that these were public Web sites, and since anyone can view and analyze them, so could we. We got some assistance from PBworks in analyzing these sites, particularly in giving us lists of the URLs of publicly-viewable, education-related wikis. They also provided us with some usage information about these wikis, such as the number of registered users, number of edits, number of new pages and so forth. Since all of this information is accessible on the Web, it would have been possible for us to devise analytic tools that would have gathered all of this information ourselves. PBworks eased our task, but they did not provide us with privileged information.

For wikis set to be privately-viewable, we had additional constraints. Our IRB required that we solicit the permission of the wiki creator, but not all members. We only attempted to get permission from wiki creators who were teachers, so we did not attempt to solicit permission to view wikis created by students (which would have required parental consent). We also suggested that wiki creators inform participants that we would be viewing, and we asked them not to invite us to join if they believed that doing so might put someone at risk of any kind of harm. We also committed to never making any kind of change to a wiki, only passively viewing.

As an additional precaution, as a group, we decided not to publish wiki URLs or direct text from wikis in our reporting. We knew from our content analysis that students did not always follow best practices in terms of protecting their identities online (though for the most part they did). Therefore, we decided to describe wikis and their activities, but not to quote content or share links that would bring additional scrutiny to the wiki. We are willing to share this data on request with researchers who would like to conduct additional analyses on our data under the guidance of their own IRB.

Overall, we found that most educators and students were very grateful for the opportunity to share their stories and lend their thoughts to our efforts to understand the use of wikis in K-12 settings. Our IRB at Harvard was very willing to work with us to develop protocols to explore this new domain. As technology develops, researchers and IRB staff will need to continue to collaborate to find safe, effective strategies for studying learning in online environments.

Section 6: Overall Ratings

In addition to the 24 items of the WQI, our coders also made four “overall quality” ratings for every wiki. They assessed the degree to which wikis provide opportunities for students to participate in the wiki, and to develop expert thinking, complex communication, and new media literacy skills. They rated these four domains on a 7-point Likert scale. Research assistants were encouraged to use any criteria that they wanted for these four overall ratings. They could consider our 24 items and/or they could consider any other evidence that might influence their assessment. We asked them to attempt to be consistent internally, but they were not required to have their criteria for the overall ratings cohere to any set of group norms.

We designed these ratings with two purposes. First, we hoped that these more subjective ratings would help us identify dimensions of quality that are not captured by the WQI. For instance, we can examine wikis with high ratings but low WQI scores, or low WQI scores but high ratings to attempt to identify the causes of these discrepancies. Second, we wanted to test whether certain questions might be answered more efficiently with overall ratings rather than quality coding. For instance, if ratings correlated highly with WQI scores, then ratings might prove to be a more time efficient method for evaluating wiki quality that the WQI.

To date, we have not conducted any analyses using the overall ratings. The costs of gathering the ratings was very low after research assistants had done the WQI coding, so we are not concerned if it takes us sometime to circle back to these data.

Section 7: Training protocols

In order to generate and maintain high levels of interrater agreement, our training regime involves three processes: an initial orientation, practice coding of a training set of wikis, and ongoing team meetings. Our goal with this training regime is to develop a cohesive understanding of the content analysis process among our entire team of research assistants. We try to only hire new research assistants once per year, so they can work as a team all year. We also attempt to only hire research assistants with classroom teaching experience, so our assistants will come to the work with a general sense of how schools and classrooms in the United State typically operate. In our most recent studies, our team has comprised approximately 10 people working 10 hours per week each.

When research assistants are brought into the team, they are given an extensive orientation to PBworks wikis and the Wiki Quality Instrument. We first ask new RAs to examine a representative set of wikis and explore this set so they get a general sense of the learning environments that we study. We also ask research assistants to read the Wiki Quality Instrument. We then have a series of introductory meetings that explain the goals of our research team, reviews our publication history, and introduces team members to our work process. Afterwards, we provide a detailed orientation to the Wiki Quality Instrument over several additional meetings. We review each demographic and quality item's decision rules, show examples of wikis that meet the criteria for each item, and allow experienced team members to discuss common difficulties. We also have experienced coders discuss the strategies that they use to do the detective work required by the demographic questions and to do the content analysis required by the quality items of the WQI. We then provide new research assistants with an introduction to the Wiki Coding Tool, including an online video, an online test of their ability to use the features of the Wiki Coding Tool, and finally an in-person meeting to resolve questions.

After this orientation process, coders begin to practice their skills in a training set. This training set is developed by experience coders who independently use the WQI to evaluate the demographic and quality items and then come together to agree upon a set of correct answers. In developing the training set, we code a large number of wikis, perhaps 200, and then purposively select 50 to be included in the training set. We try to include in this set several kinds of wikis. First, we include wikis with difficult to find information, or information that can only be found if coders use some of the strategies we have developed for systematically analyzing wikis. Second, we include typical wikis that have usage patterns commonly found within the set. Finally, we include wikis that have items that proved difficult to code and difficult for even our experienced raters to agree upon.

We then give coders the first 25 wikis to evaluate at each appropriate occasion of measurement, so a typical training set will include 75 or 100 rows. Research assistants are required to reach 85% agreement with the training set across all categories of the WQI and have an average composite wiki-quality score that falls within 1.5 points of the agreed correct average scores before being allowed to begin coding new wikis. After all trainees have completed their first 25 training wikis, we conduct an analysis of the trainee scores to determine which WQI categories have the most disagreement and which wikis have the most disagreement. We then hold a meeting where we give trainees the correct scores and share with them our analysis of disagreement. We review wikis and categories that were problematic, and we answer questions. Trainees are encouraged to go back to the original 25 wikis to review and correct their errors.

If trainees meet acceptable levels of agreement on the first training set, we have them start coding new wikis. If not, they are given the second training set of 25 additional wikis. If they are successful after the second set, we have them start coding new wikis. If not, by this point our group has conducted analyses of many additional wikis that we can use to create additional training sets. In our experience, we could get all of our trainees at acceptable levels of agreement within four training rounds. If we were not successful, we would have found other research tasks for that individual or counseled him or her out of our research group.

This extensive process takes weeks, and the team that we assembled in August was not functioning at full capacity until October.

To maintain a close alignment of scores, research assistants participated in weekly meetings to discuss wikis and quality categories that were particularly difficult to code. In the early months of the year, our training discussions were mostly driven by questions that new research assistants had about difficult wikis to code. As the year went on and we had more data from reconciling disagreements, we could target our discussions to focus on areas where we knew we had low levels of disagreement. We would often collectively analyze difficult wikis.

We also frequently asked research assistants to revisit the WQI. As we shifted from demographic coding to quality coding, or as we shifted from one sample of wikis to another, we required that our research assistants re-read the WQI and revisit the decision rule language that is fundamental to our ability to code in agreement.

Section 8: Conclusion

We believe that the availability of real-time data from online learning environments represents a watershed moment in education research. These data allow researchers to examine detailed records of student-teacher interactions in depth and at scale. Over the past three years, our group has put significant resources behind developing methods for content analysis in diverse learning environments, and our goal in publishing these methods online is to support other researchers who are attempting similar projects.

We welcome questions and feedback from fellow researchers and other interested parties. While we have chosen not to publish here all of the specific worksheets that we used or training sets that we developed, we are happy to share these materials with other researchers with a material interest in those parts of our research program.

Part III: On the Development of the WQI

We developed the Wiki Quality Instrument (WQI) to answer two kinds of questions about the role of wikis in K-12 education: 1) How do educators design wiki learning environments that promote rich learning experiences? 2) Do only certain learners have access to these high quality wikis? To answer these related questions, we needed a means to evaluate the degree to which wikis support high quality learning. No such measurement existed at the beginning of our research process, so designing our own instrument has been a signature feature of our research agenda.

The WQI is a content analysis rubric used by trained research assistants to evaluate U.S., K-12 wiki learning environments. The WQI is intended to be used with a large sample of wikis, where each wiki is measured at multiple occasions. We use these multiple quality measures to produce longitudinal wiki-quality profiles. These quality profiles represent wiki quality as a trajectory rather than as a single measure. Since our goal is to conduct our investigations at a scale of hundreds or thousands of wikis, we designed our WQI to require approximately 30 minutes, on average, to conduct one evaluation of one wiki at one time point.[5]

In this document, we describe the process by which we designed the Wiki Quality Instrument. First, we set some context for our study by describing the unit of analysis in our study and the scale and scope of our inquiry. These contexts provided a variety of constraints to our instrument design. Next, we summarize the research we used to develop the theoretical framework of the WQI. To determine the domains of wiki quality,

---

[5] See Part II for analyses of actual coding times; 30 minutes was our target.

we examined the literatures on 21st century skills and on evaluation of online learning environments. We also gathered data from wiki-using teachers and students through surveys, interviews, and classroom observations. We explain here how we synthesized those perspectives into a theoretical framework. Finally, we describe how we used an iterative process to operationalize our theoretical framework into a set of valid and reliable items to create the WQI.

Defining the Context of our Study

Before delving into the details of our instrument design process, we highlight two important features of the context of our research. First we define the wiki subdomain as our unit of analysis. Second, we explain the advantages and constraints associated with the scale of our inquiry.

What do we mean by "wiki"? The wiki subdomain as the unit of analysis:

The unit of analysis in our study is the wiki subdomain. A wiki subdomain is a particular Web address provided by a wiki hosting service. For instance, PBworks—the wiki hosting service that supplies the wiki data for our project—uses the domain "PBworks.com" and allows users to create subdomains, such as ReichWorldHistory2009.PBworks.com. We use these subdomains to draw rigid conceptual boundaries between wiki learning environments. Thus we consider ReichWorldHistory2009.PBworks.com as one "wiki" and ReichWorldHistory2010.PBworks.com as a second "wiki." Much of the material on those two wikis might be the same, but in all likelihood the students would be different;

the project endured even as the students changed. It might be that ReichMedievalProject2009.PBworks.com and ReichRennaissanceProject2009.PBworks.com are different projects completed by the same classrooms of students. Reich2009TeamA.PBworks.com and Reich2009TeamB.PBworks.com might be from the same class and doing the same project. All of these, we would define as separate wikis in our analysis, even though in other kinds of studies sensible researchers might choose to treat all of these different subdomains as one “wiki community”.

By using the wiki subdomain as our unit of analysis, we could apply clear, automated decision rules to defining wiki communities. Each subdomain was treated as a separate, unique case in our dataset. Choosing the wiki subdomain as our unit of analysis also has certain technical benefits. For instance, PBworks maintains their usage statistics at the subdomain level. Moreover, from our preliminary analysis we observed that relatively few wiki subdomains appeared to be nodes in the kinds of networks of wiki communities hypothesized above. Most users create wiki subdomains as discrete entities.

To be sure, in other kinds of studies, it might be quite profitable to attempt to link together related subdomains. For instance, in a study closely examining wiki usage in particular school settings, researchers might study a school where wikis are used frequently and wiki projects are connected to one another in meaningful ways. While it might be somewhat technically difficult to track discrete users as they make contributions across multiple subdomains, such an effort might be worthwhile in very closely examining a particular group of wiki users. Such an effort, however, would be very difficult and very expensive to do at scale. For our purposes defining a wiki as a wiki subdomain was a superior approach.

Henceforth, when we refer to a “wiki” in our dataset, we are referring to a publicly-viewable, education-related wiki subdomain hosted by PBworks.com.

What is the scale of inquiry? Managing wiki study at scale

At the heart of our research agenda is the belief that there is something to be learned from quantifying characteristics of the entire universe of U.S., K-12 wikis. To gather a representative sample of this population, we chose to study samples that include hundreds or thousands of wikis. This led to three challenges that put constraints on how we define quality and how we develop our coding protocols. These three challenges are 1) the diversity of the universe of wikis, 2) the scale of our investigation, 3) the inability to track individual users with acceptable levels of reliability.

In our preliminary analysis, we found an extraordinary diversity of activity in the universe of wikis. In our sample of 1,799 wikis drawn at random from 179,851 publicly-viewable, education-related wikis hosted by PBworks, we identified wikis used in elementary schools and wikis used by seniors in honors classes. We found wikis supporting instruction in virtually every academic subject area: English/language arts, social studies, math, science, computers and technology, foreign languages, arts, and physical education. Wikis are also very flexible platforms, so teachers and students used wikis in manifold ways: as online handouts, online worksheets, platforms for collaborative presentations, discussion forums, topical encyclopedias, and student portfolios.

All of this diversity presents serious challenges for designing a measurement instrument that can be used reliably by a team of coders. In order to manage the diversity we faced in student ages and levels, we did not define quality in regards to particular details of student performance. Rather, we more broadly looked at the kinds of learning opportunities that students had in wiki learning environments. For instance, we documented seven different types of student collaboration, such as copyediting. We did not measure the efficacy of the particular discursive moves made by students in collaboration with each other, such as measuring the degree to which a copyeditor made focused, constructive suggestions. We can reliably identify when a third grader is copyediting another student's work as easily as we can reliably identify when a senior in high school is copyediting. Attempting to measure the specific quality of the copyediting activity would have been too complex for the scale of our inquiry.

Similarly, in determining what kinds of learning opportunities to evaluate, we only chose to evaluate opportunities that would be common across the academic subject areas. If we were to, in a future study, constrain our inquiry to a single subject domain, such as science, then we would have more opportunities to choose specific indicators that would give a richer indication of the quality of science instruction and learning occurring on the wiki. If we refined yet further—to Earth science, or 7th grade Earth science taught in Iowa—then we could be even more detailed and specific in our criteria. We believe, however, that at this stage in the development of research into Web 2.0 tools and deeper learning in K-12 schools, we need a broad national perspective. Thus, we have eschewed specificity and sacrificed some measure of depth in order to maintain this broad view.

In addition to the challenges presented by the diversity of the universe of wiki activity, we also faced the challenge of the scale of our inquiry. Because we attempted to manually code wikis at considerable scale, we quickly confronted constraints of time and resources. When we were developing our instrument, we made the following assumptions about the costs of applying the instrument. We wanted to code 500 wikis on 4 separate occasions, which would require conducting 2000 evaluations. Then, we also wanted each wiki to be evaluated twice by independent raters, so we needed to conduct 4000 evaluations. If each evaluation were to take 30 minutes, and if research assistants were to bill at $12/hour, then the cost of applying the WQI would be approximately $24,000. When we added the costs of training coders, evaluating coders, holding meetings to maintain consistency, and reconciling discrepancies, we assumed that the costs would exceed $30,000. Using these estimates, for every additional minute that it would take a research assistant, on average, to code a wiki, the cost of our pilot investigation would increase by approximately $1,500. Of course, not all these assumptions proved to be true. Federal Work-Study grants reduced our labor costs; in our first sample we coded approximately 400 instead of 500 wikis; we used six occasions of measurement instead of four, but since wiki lifetimes are so short, most wikis required three or fewer measurement. The point here, however, is that to examine wikis at a scale of hundreds or thousands, we needed an instrument that could be applied relatively quickly.

The scale of our investigation, therefore, set constraints on how deeply coders could evaluate each wiki. Early on, we recognized that any efforts at scalar comparison were unlikely to be successful. In order to design the WQI to be used in a 30 minute evaluation, we chose quality indicators that were relatively simple to detect and could be measured and evaluated without extensive textual analysis or calculation. We chose to evaluate the presence or absence of different behavior types rather than the frequency of particular behaviors or the quality of particular behaviors.

For instance, in our preliminary analysis, we found that coders could achieve acceptable reliability when coding for the presence or absence of copyediting. Coding for the frequency of copyediting on a particular wiki could not be accomplished reliably in a reasonable length of time. The variation in wiki size was a major contributor to this dilemma. Consider two wikis. One wiki has one page, and that page is copy-edited in a few places. Another wiki has over 100 pages and only two exhibit copyediting. One of those pages is copyedited extensively, the other only barely. Even with these starkly contrasting scenarios, it is not difficult to imagine the dilemmas caused by trying to create a single, unified scale measuring the frequency of copyediting. Moreover, as noted above, we had very little hope of measuring the quality of copyediting across very diverse wikis. Thus, the WQI measures the presence and absence of behaviors that promote high quality learning rather than the frequency or degree of those behaviors.

Our third challenge was the difficulty of tracking individual users. From our observations and content analysis, we knew that users do not always conduct all of their wiki activity using a unique login. For instance, in one third-grade classroom in San Diego, we observed a teacher who created a wiki, signed in each day with his own login, and then allowed students to take turns contributing under his username. A content analysis of the wiki would easily reveal that this wiki is co-constructed by third-graders, but we have no way of knowing exactly which third graders are responsible for which contributions.

This uncertainty curtails our ability to measure quality by measuring changes in individual student behavior or performance, which of course is one of the most important indicators of quality in classrooms. In future studies, we hope to partner with schools or districts using online learning environments to evaluate the quality of 21st century skill development in online learning environments by tracking the performance of individual students. In our circumstances, however, we chose to forgo efforts to track individual student development in order to evaluate wiki usage and quality at a national scale.

These challenges of wiki diversity, the scale of our inquiry, and the difficulty of tracking individual users shaped our study in several fundamental ways. In order to deal with these constraints we developed an instrument that focused on broadly applicable markers of quality, that evaluated evidence of opportunities for learning rather than evidence of measurable cognitive improvement, and that measured the presence or absence of types of learning opportunities rather than frequency or the level of quality inherent in those types of learning opportunities. These constraints were integral to our thinking as we developed our theory of wiki quality and the WQI.

Developing a Theory of Wiki Quality

Our process for developing a theory of wiki quality involved three kinds of research. First, we were committed to listening to the voices of teachers and students in our instrument design process, so we used several methods to listen to, record and analyze their experiences. Through teacher surveys, teacher interviews, student focus groups, and classroom observations, we explored the ways in which teachers and students defined and assessed wiki quality. Second, we wanted to build upon any relevant published scholarship, so we conducted a wide-ranging literature review to examine how previous researchers had evaluated quality in online learning environments. Third, we conducted an additional literature review on the theme of 21st century skills to examine how educators, researchers, and policymakers conceptualized high-quality learning beyond the specific domains of online learning environments. From these three analyses, we developed a theoretical framework for measuring wiki quality.

Before delving into these research methods, it is important to highlight one of the assumptions that we brought with us into the research. Any effort to measure wiki quality at a national scale assumes that certain dimensions of quality are universal across American education. One reasonable position on educational quality might posit that high-quality learning environments are those that meet the learning goals established by the students and educators within that community. From that perspective, universal measurement is folly, since the only meaningful markers of quality are those that are locally defined. While we have some sympathy for the position that quality teaching responds to local contexts, we reject the notion that dimensions of quality are entirely defined and contained locally. In order to measure wiki quality at a national scale using our methods and resources, we could not assess wikis in their classroom contexts. We could study 500 wikis, but we could not study 500 wikis and their 500 associated classrooms across the country. From the beginning, therefore, we resolved to study wikis divorced from their larger learning ecology. We assumed we could identify certain universal features of “good” wikis without knowing specific contextual details from these larger learning ecologies. This is not to say that this approach is “better” than research studying smaller numbers of learning environments, but our approach offers a different perspective, a different trade-off between depth and breadth.

How does the Literature on 21st Century Skills Define High Quality Learning

Nearly everyone who studies education for a living has a set of broad assumptions about what makes for high quality learning environments. Within our research team, these assumptions were strongly influenced by the work of Frank Levy and Richard Murnane. In their book *The New Division of Labor*, Levy and Murnane (2004) used labor market research to develop a taxonomy of skills critical for success in 21st century job markets. They argue that computers have taken over a considerable portion of routine manual and cognitive tasks in the workplace, and thus the labor markets have shifted to include more jobs requiring skills that computers cannot perform well: expert thinking and complex communication. Expert thinking is required for solving ill-structured problems, tasks that cannot be completed with rules-based logic and tasks requiring tacit knowledge. Complex communication is required for tasks which are defined or accomplished through social interactions. Thus, in Levy and Murnane's formulation, expert thinking and complex communication are 21st century skills in the sense that they are of growing importance in 21st century labor markets. Their research is a cornerstone of empirical research on what are now known as 21st century skills, and it is important for us to acknowledge that this research shaped our thinking entering the project. That said, we made every effort to remain open to alternative perspectives from teachers, students, or the published record of research as we developed the WQI.

One of our first research steps, therefore, was to examine other research defining the skills, knowledge, and competencies that educators should value. Since the publication of Levy and Murnane's work, many other researchers, thinkers and policymakers have developed other lists of 21st century skills (Gardner, 2006; Haste, 2008; H. Jenkins, 2006; Partnership for 21st Century Skills, 2007; Trilling, Fadel, & Partnership for 21st Century Skills, 2009; Wagner, 2008). We examined these lists, and we developed matrices for comparison among the different frameworks. Eventually, we came to conclusions similar to Dede's (2010) analysis of several of the most prominent frameworks of key skills for the 21st century. He found that expert thinking and complex communication are featured in nearly all of the well-regarded 21st century skill lists, along with one other domain: technological literacy. This seemed to be a domain well aligned with our interests in wikis.

To develop our own formulation of technology literacy, we borrowed from Jenkins' (2009) definition of new media literacy. Jenkins argues that emerging networked technologies require that students have the ability to critically consume and produce diverse forms of social media in a collaborative, networked context. In a sense, this definition of new media literacy defines a particular sub-category of complex communication, a category which includes tasks which are defined and accomplished by communicating with diverse forms of multimodal media. Thus, from early on in our research, these three domains—*expert thinking, complex communication, and technological literacy*—formed the core of our working definition of 21st century skills.

While we were doing our own reading and thinking about the key elements of high quality learning, we were simultaneously working to bring the voices of teachers and students into our deliberations. In the next section, we discuss how we gathered data from classroom wiki users and what they had to say about wiki quality.

How do Wiki-Using Teachers and Students Define and Assess Wiki-Quality?

We used multiple methods to investigate how teachers and students in wiki-using classrooms defined and assessed high-quality work in wiki learning environments. We interviewed 68 teachers from across the country about their use of wikis. Approximately half of these subjects were randomly drawn from our sample of 411 U.S. K-12 wikis, and the rest were purposively recruited as expert wiki users (recruited through personal contacts), teachers in urban schools (recruited primarily through personal contacts), and novice wiki users (recruited from a PBworks summer institute for teachers). We also visited 19 classrooms in six U.S. states. In sampling these teachers, again we used a combination of cold-calling randomly selected wiki creators identified on large lists of wikis and contacting particular teachers through personal contacts. From these diverse sampling efforts, we believe we captured a broad cross-section of wiki users. While visiting classrooms, we also recruited students to participate in focus groups with our researchers, and we conducted over 40 student focus groups through these methods. We also surveyed 192 participants in an online wiki summer professional development program designed for novice wiki users and hosted by PBworks, and we asked them what they anticipated to be the benefits of using wikis with their students. We compiled field notes, interview transcripts, and analytic memos into an electronic qualitative research package, and we analyzed our data looking for common themes voiced by teachers and students.

Through these various channels, teachers expressed a diverse set of beliefs about the benefits and affordances of wikis. When we asked teachers why they chose to invest their time and energy into developing wiki learning environments, they discussed several major categories of benefits. Wikis gave students opportunities to develop communication and collaboration skills, ranging from commenting on each other's work, to peer editing, to co-creating projects and assignments. Wikis allowed students to develop a fluency with a new technology platform, and to publish multimedia presentations of their arguments and beliefs. Wikis also simplified some of the logistics of classroom communication. They allowed teachers and students to share both logistical information about the course such as homework assignments and classroom guidelines as well as materials related to course content. Teachers also viewed wikis as places for students to develop and publish projects, and to deepen and display their understanding of course skills and knowledge. When discussing what makes a great wiki, advanced wiki-using teachers described sophisticated performances of understanding where students demonstrated mastery of course content, collaboration skills, and technological design competencies.

The reasons that teachers described for using wikis overlapped with our own understanding of high-quality learning environments seen through the lenses of complex communication, new media literacy, and expert thinking. Teachers' discussion of peer editing and collaborative projects fit into the domain of complex communication. Their comments about technology literacy and using the multimedia affordances of wikis cohered with our ideas about new media literacy. Their descriptions of project-based work, of self-directed work, and of developing and displaying understanding cohered with our domain of expert thinking. From analyzing teacher's descriptions of why they used wikis, we felt that their purposes aligned quite well with our conception of 21st century skills.

The new domain that teachers introduced to us had to do with the logistics of classroom life. They described the importance of having a central place to post course materials, course content, and links to related websites. They also described the importance of having a place where students could post questions, homework, and links to other materials and interact with the class outside of class time. We defined this domain slightly more broadly than logistics, and we began to refer to these practices as elements of *participation*. We viewed these basic ways of interaction—reading materials, following links, posting simple content—as the precursors of more sophisticated behaviors that promoted deeper learning.

Overall, we felt that our theoretical conceptions of high quality learning environments cohered well with the ideals and objectives described by our wiki-using teachers. The domains of expert thinking, complex communication, and new media literacy resonated with their descriptions of the benefits of wikis and of their goals with integrating wikis into their classroom. From their descriptions, we also resolved to add the conceptual category of *participation* to our analytic framework. Since the degree of alignment between teacher values and our beliefs about wiki quality was quite high, we probed the data further to see if teachers had assessment mechanisms that might help us develop the items of the WQI.

Along with asking teachers to talk about what they valued, we also asked teachers and their students to describe how teachers measured and assessed quality work in wikis. In our interviews, we often prompted teachers to discuss both their formal assessment mechanisms—like grades and rubrics—as well as other informal mechanisms, like comments they might leave on a wiki or share with students in class. We asked students about what they thought good work looked like on wikis, who they thought was doing good work, and how their teachers graded and evaluated their efforts. We hoped that some of these assessment mechanisms might prove useful in developing the WQI.

What we discovered was a striking disjunction between what teachers said they valued and what they actually graded. Our evaluation of the overall trend of teacher assessment in wiki learning environments is that teachers primarily grade students for "following directions." Many teachers reported that they evaluate students for participating in the wiki community at prescribed intervals, for including the required number of design elements (like pages, paragraphs or images), and for including factually correct content. These routine tasks generally do not cohere with the domains of 21st century skills. Some teachers did report assessing students' communication, collaboration, and technology fluency, and a few reported assessing understanding or critical thinking. Many teachers reported that they valued the skills in these deeper learning domains, but they struggled to figure out exactly how to assess deeper levels of understanding or expert thinking. One teacher referred to this as a "gray area" in grading, and explained that while he certainly valued these domains, he was not sure how to develop objective assessment criteria. This teacher's dilemma is not surprising, as developing strategies for assessing 21st century skills and higher-order thinking skills is an unsolved challenge at the heart of numerous research efforts across the world.

We did not discover any commonly used assessment criteria for expert thinking, complex communication, or new media literacy that were adopted across multiple classrooms in our study. In developing the WQI, we used our qualitative data as one source of data about the various discursive practices that occur on wikis and as a source of inspiration of WQI items. We did not, however, find -any common grading or evaluation practices that we could directly modify or adapt in creating our WQI. While searching these data for these kinds of metrics of 21st century skill development, we also turned to the research literature on measuring quality in online learning environments.

How have other scholars approached measuring quality in online learning environments?

In developing our definition of quality and our WQI, we conducted an extensive literature review to investigate how other researchers and scholars had approached the evaluation of Web 2.0 learning environments. We conducted searches for terms such as wiki*, blog*, and “Web 2.0” in databases of published articles and unpublished theses. We also examined all articles in the last ten years of the *Journal of Learning Sciences*, the *International Journal of Computer-Supported Collaborative Learning,* the *Journal of Research in Technology in Education*, the *Journal of Computer-Assisted Learning*, and the *American Journal of Distance Education*. Research assistants read and summarized all of the articles in these volumes that dealt with measuring quality or with Web 2.0 tools in educational contexts.

Research into Web 2.0 learning environments—wikis, blogs, discussion forums, proprietary environments, and other platforms—has primarily been conducted through small-scale design research experiments and qualitative case studies. Most studies examine one or a few classes of students, often in courses taught by the researchers. These studies typically investigated a single narrow dimension of student learning, such as cognitive engagement (Oriogun, Ravenscroft, & Cook, 2005), collaboration (Cortez, Nussbaum, Woywood, & Aravena, 2009; Trentin, 2009), or the effect of incongruity between knowledge and information on knowledge building (Moskaliuk, Kimmerle, & Cress, 2009). Often these studies were conducted within a single subject domain, such as algebra (Chiu, 2008), business ethics (Jeong, 2003), or American history (Lawrence, 2009). The studies used a wide variety of methods to assess the quality of learning environments and student development, including pre-post student testing, pre-post student surveys, and content analysis of online materials.

Our study took a significant departure from these approaches. We studied samples of hundreds of wikis drawn from populations of hundreds of thousands of “naturally occurring” wikis rather than examining special sites or our own classrooms. As a result, we studied learning environments naturalistically by examining the work that teachers and students were already doing, rather than devising interventions or design experiments where conditions were controlled to test particular hypotheses. We studied wikis that support instruction across all the subject areas rather than just in one particular academic domain. We sought to evaluate wiki quality broadly rather than one specific dimension of quality.

Perhaps the study that comes closest to ours in scope is Kozma’s (2003) work analyzing 174 case studies of innovative technology projects identified in 28 countries. Kozma assembled an international team of researchers. They created a list of exemplary education technology projects from each country, constructed case studies for each project from interviews, observations, and document analysis, and then classified the case studies in a number of categories. The fundamental similarity between our work and Kozma's research is that we both examine evidence from technology-based learning environments, classify that evidence, and then conduct quantitative analysis to better understand how teachers develop rich, technology-based learning environments. Sufficient differences between our projects, however, limit the compatibility of Kozma's research tools with our approach. Kozma's unit of analysis was the case study, which included evidence both from the technology application and the classroom context. Our unit of analysis was the wiki itself, and we limited our examination to the technology platform. Kozma studied only exceptional cases; we studied the full distribution of wiki learning environments.

Both focused and broad studies of Web 2.0 learning environments are much needed. Up to this point, however, most research into Web 2.0 learning environments has been conducted by examining particular learning environments under microscopes, and in this study we attempted to characterize the universe of wiki learning through telescopes. Previous researchers have developed many innovative approaches to studying online learning environments, but most studies had a different grain size than our own investigation. The fine-grained measurement mechanisms that most researchers used to evaluate a particular dimension of quality in a particular domain seemed inapplicable to our efforts to more broadly evaluate quality in a diverse universe of K-12 wikis. As a result, while we could draw some parallels between previous quality measurement approaches and our own efforts, after an extensive review of existing approaches to measuring quality in online learning environments, we chose not to directly adapt existing instruments for evaluating wiki quality into our own WQI.

Iterating towards the Final Wiki Quality Instrument

After our qualitative research and literature review, we had established four domains of wiki quality with backing from the research literature and alignment with the objectives of wiki-using teachers: participation, expert thinking, complex communication, and new media literacy. Our next challenge was operationalizing these domains into items that we could use to create a content analysis rubric.

Developing these items was an iterative process that took place over a year. Our first efforts took a grounded theory (Charmaz, 2006) approach, where we conducted several rounds of open coding of wikis to get a sense of the kinds of behaviors that we could identify on wikis. We then did multiple rounds of focused coding where we tested a variety of item types. When then conducted additional rounds of pilot coding in order to conduct interrater agreement analysis and finalize our items. The final step in revising the WQI came towards the end of our first study, when we had sufficient data from our quality measures to conduct principal components analysis and cluster analysis in order to assess coherence of the items within our domains.

In parallel with our literature reviews and qualitative research, we also analyzed wikis directly. Many of our early rounds of wiki coding were focused on identifying basic demographic features of the wikis. For instance, we sought to separate out U.S., K-12 wikis from wikis used in other countries and in higher education. Once we established which wikis were used in U.S., K-12 schools, we sought to classify them by subject area, by grade level, and by their hosting institution (school, library, district, etc.). In each of these early rounds of demographic coding, we also did various exercises to analyze the content, teacher activity, and student activity of the wikis.

For instance, in our first round of wiki coding, we asked research assistants to briefly describe the "purpose" of each wiki. We did not provide criteria or guidelines for the exercise, though we asked coders to try to settle on internally consistent language within the set of wikis they analyzed. In other words, if they started using the term "student portfolio" to describe a subset of wikis, we asked them to individually work out a set of decision rules for applying that term consistently. From these qualitative descriptions, we collaboratively developed a taxonomy of wiki purposes. This taxonomy included categories like "trial wiki," "individual student, single assignment," "individual student, project," "individual student, portfolio," "collaborative student, single assignment" and so forth. In subsequent rounds of wiki coding, we asked raters to attempt to classify the wikis by our agreed upon purpose categories. "Purpose" proved to be too nebulous a concept and we could not generate sufficient interrater agreement in our classifications (itself a useful finding), but the exercise did give us a better sense of the kinds of things that teachers and students did with wikis.

While refining our purpose categories, we also asked research assistants to describe "patterns of practice" they encountered on the wikis. These patterns of practice were identifiable discursive moves made by teachers and students to facilitate student learning. Again, we gave coders very few guidelines for what might constitute these patterns of practice. We did ask them to think about our four conceptual quality categories, of participation, expert thinking, complex communication, and new media literacy. Beyond that, however, we asked them to simply write about what they saw happening. At this stage, we examined over 400 wikis with two raters looking at each wiki, so we developed a pretty extensive set of qualitative descriptions of wiki activities.

We also, in these early rounds, began testing preliminary items. For instance, we developed a four-item taxonomy of behaviors displaying complex communication: concatenation, copyediting, co-construction, and commenting. We considered whether we could attempt to create some kind of quality scale for these items, but we realized that it would be impossible to quickly and reliably assess “good” copyediting versus “bad’ copyediting. We did attempt to make a simple, scalar assessment of the frequency of these activities by using a 0-2 scale where 0 was “activity not found,” 1 was “activity found infrequently” and 2 was “activity found regularly.” We did not provide precise definitions for the frequency categories. We found that we were unsuccessful at reliably rating the frequency of these four collaborative activities, but we were successful at reliably identifying the presence or absence of these activities. Moreover, wikis with evidence of multiple collaborative characteristics did appear to be generally more collaborative than wikis with just one characteristic. We also discovered that certain behaviors, such as signing up for a timeslot or a responsibility on a list, did not fit well within our complex communication schema. So in future iterations of the WQI we added items for planning, scheduling, and discussion.

Through additional rounds of pilot testing, we attempted several other approaches towards item design. For instance, we developed a set of indicators of technology use for our new media literacy category. These items included using formatting, adding links, and embedding images. For a while, we tried to distinguish between “substantive” and “decorative” uses of these elements. For instance, when did formatting really enhance the argument or artistic message of a wiki page, and when was it simply meaningless decoration? This was another effort at scalar measurement, and once again we could achieve agreement on the presence or absence of formatting, but we could not reliably distinguish decoration from substantive uses in a timely fashion.

In another pilot version of the WQI, we tried to identify both the presence of an activity as well as the intention for the activity to take place. In some wikis teachers indicate that certain behaviors are supposed to happen. For instance, a teacher might assign students to comment on each other's work. We attempted to measure both when an activity actually happened and when a teacher intended for the activity to happen. Measuring teacher intent, however, quickly devolved into an exercise in parsing and mind-reading with low reliability, and we abandoned the effort.

While refining the item categories, we also refined our decision rules for each item. We found early on that long decision rules that listed many examples of the presence and absence of a behavior led to disagreement. When decision rules listed many specific examples, some coders only looked for those examples while others looked for the general principle. Based on this experience, for each item we wrote relatively short decision rules that focused on the general principle without many examples. We also experimented with phrasing our decision rules as questions, but we found it more effective to define decision rules as pairs of declarative statements describing the presence and the absence of the behavior. We still use the "question format" in publications as a summary of our instrument, but coders do not use the questions.

Thus, through numerous rounds of pilot testing, refinement, and iteration, we settled upon a near-final version of the WQI. In our last round of pilot coding, before we began training a new set of research assistants, we had two senior research assistants code a set of new wikis with the instrument. Afterwards, they sat down to discuss their disagreements, and we used these points of disagreement to make additional refinements to our decision rules. We also used some of these difficult wikis in our training set for new research assistants, to give them a sense of some of the challenges of coding wikis consistently.

When we started the first round of wiki coding, we had 25 items in four subdomains. There were two differences between that version of the WQI and the one that we reported in our early publications. In the original specification of the WQI, the participation subdomain included six items: Course Materials, Information Gateway, Contribution, Individual Page, Shared Page, and Student Ownership. In the complex communication subdomain, the WQI included the present seven items as well as one item for Beyond Classroom Communication, which evaluated whether students from more than a single classroom interacted on the wiki. We changed these items after coding the wikis for our first study and using principal components analysis to determine if our theorized subdomains in fact clustered together.

We made two changes to the instrument based on our cluster analysis. First, we deleted the item concerning Beyond Classroom Collaboration. This behavior was so rare, that the item artificially inflated our overall interrater agreement (it is easy to agree that something that never happens), and it did not cohere well with the other items in the complex communication category. We also separated out the Course Materials and Information Gateway categories out of the participation subdomain. Theoretically, the reason to include them in the participation subdomain was that they represented basic ways for students to interact with the wiki. However, since many wikis consisted of *only* students engaging with the wiki through view course materials and links, principal components analysis showed that wikis with positive scores for these two categories tended to score a 0 in all other categories. As a result, we created a fifth subdomain, *Information Consumption* , based on our empirical data, which included our two items for Course Materials and Information Gateway

At this point, we expect the current version of the WQI, with 24 items in five subdomains, to remain stable as we continue our data analysis on additional wiki samples.

Summarizing the Design Process for the Wiki Quality Instrument

Our process of instrument design included six major steps.

1) Defining a theoretical framework for wiki quality based on the literature regarding 21st century skills
2) Conducting qualitative research with wiki-using teachers and students to determine how they defined and assessed wiki quality
3) Conducting a literature review of efforts to measure quality in online learning environments in order to assess whether existing items, scales or instruments could be integrated or adapted for our purposes
4) Conducting several rounds of open coding on wiki learning environments in order to develop a taxonomy of common patterns of practice on wikis

5) Conduct multiple rounds of pilot testing to test different items, scales, and decision rules
6) After data collection and analysis, make final revisions to the instrument based on cluster and principal components analysis.

Designing this instrument has been a balancing act. On the one hand, we sought to identify important indicators of potential opportunities for 21st century skill development. On the other hand, in order to investigate wikis at scale, we have ensured that the indicators we chose to examine can be evaluated reliably and relatively quickly. This WQI was designed to be used in a research program where we make thousands of evaluations by examining hundreds of wikis on multiple occasions. Also, it is designed to be used in evaluating a very diverse population of wiki learning environments from all subjects and grade levels. We believe that this foundational instrument can be refined and improved to be even more useful, valid, and nuanced in evaluating more specific subpopulations of wiki learning environments.

Part IV: On Developing Protocols for the WQI

From the beginning of our study, we have taken the position that quality is a time-varying feature of online-learning environments. Since wikis preserve their entire edit history in continuous time, data from these wikis are uniquely well suited to evaluating changes in wiki quality over time. In theory, researchers could take quality measures every second on a given wiki and reconstruct wiki quality in continuous time. However, since our WQI instrument requires human raters to evaluate quality, there are non-trivial costs to taking quality measures (see Part II on Content Analysis). Therefore, in our research design, we needed to develop a data-collection protocol that balanced our desire for adequate data with the constraints of our budget.. At present, we measure wiki quality on days 1, 7, 14, 30, 60, 100 and 400. In this section, we describe the preliminary research that we conducted to arrive at these particular occasions of measurement.

In developing protocols for administering the WQI, we faced a tension between the desire to obtain measures on many occasions in order to model our dependent variable accurately and the constraints of time and money. This tension led us to address two questions about the frequency and timing of wiki-quality measurements as we created protocols for administering the WQI:

1) **Frequency:** How often should we measure wiki quality?
2) **Timing:** When in a wiki's lifecycle should we take quality measurements?

The optimal frequency of wiki-quality measurement is proportional to the complexity of wiki-quality developmental trajectories. If wiki quality develops following a linear trajectory, then three data points may be sufficient to model these linear trajectories. However, if typical wiki-quality developmental trajectories are more complex, then measurements on more occasions are necessary to model this development accurately. Thus, we needed to assess the complexity of wiki-quality trajectories in order to ascertain the optimal frequency of data collection.

While answering the question of how frequently to measure, we also needed to determine the optimal timing of those measurements. As we estimated the complexity of wiki-quality developmental trajectories, we also needed to model wiki lifetimes to determine when within the wiki lifecycle we should be taking quality measures. For instance, if wikis typically remain active and changing for 10 days, then we would have a different timing for taking measurements than if they typically remain active for 10 months or 10 years.

To address these two questions we used two longitudinal analytical methods. To assess the complexity of wiki-quality development, we used empirical growth modeling. To model typical wiki lifetimes, we used continuous-time survival analysis. We combined insights from the findings of both of these methods in order to settle on our protocol of WQI measurements. In the following sections, we summarize the design, findings, and implications of our studies using these two methods, and then we detail the reasoning that led us to our current protocol for wiki-quality measurements.

How Often Should We Measure Wiki Quality?

We faced a catch-22 as we approached the question concerning the appropriate frequency of wiki-quality measurements. Ideally, we would have made our determination of the appropriate frequency of wiki-quality measurements by evaluating the complexity of wiki-quality trajectories. It was not possible, however, to estimate wiki-quality trajectories before developing an instrument and protocol for measuring wiki quality.

We resolved this dilemma in the earliest phases of our research by using a measurement of wiki usage as a proxy for wiki-quality development. We measured the number of *page edits* (both revisions to existing pages and new page creations) to a wiki, which accounted for both page edits and new page creations. In other words, in order to create a protocol for measuring wiki-quality trajectories, we assessed the complexity of wiki developmental trajectories, using page edits as the metric for wiki development. We chose page edits for both practical and theoretical reasons. First, it seemed reasonable to us that periods of high volumes of edits and activity might co-occur with quality development. Second, PBworks provided us with data files containing page edits for a random sample if 1,799 publicly-viewable education-related wikis. (See Part II, Section 2 for details on our samples). These data files were reformatted into a project-period dataset, where each row in the dataset corresponded to one day of activity for one wiki. Thus, we had ready access to the data necessary to evaluate wiki development in terms of page edits.

We used both exploratory empirical growth analysis and statistical modeling to assess the shape of typical wiki developmental trajectories. First, we examined a series of empirical growth plots for each wiki, and then we fit a series of non-parametric local regression (loess) models to summarize the trajectory empirically. In the absence of any theoretical assumptions about the shape of wiki-quality trajectories, these approaches allowed us to assess and then model wiki development without imposing any parametric constraints on our model.

As recommended by Singer and Willett (Singer & Willett, 2003), before we did any formal modeling of wiki development, we examined a series of empirical growth plots from sample of 411 U.S., K-12 wikis. First, we created a simple scatter plot for each of several wikis with days on the x-axis and page edits on the y-axis. We also fit a simple OLS regression line to each pointcloud to approximate wiki development. After examining wikis with time binned into days, we decided that we might be attaching too much weight to what was essentially stochastic variation from day to day. Therefore, we recoded the dataset so that time was binned into months. Especially for longer-lived wikis, inspecting these empirical growth plots helped capture larger trends.

In Figure 1, we present four of these empirical growth plots. Notice that for the top two plots, wikis are more active in the early months and then become less active. For the bottom two plots, wiki activity maintains very low levels throughout the entire history of the wiki. The slopes for most of the OLS trajectories fitted to these plots were either negative or flat near zero. These patterns alerted us to two common patterns of wiki development: 1) wikis with early activity that declines gradually and 2)wikis that have little or no activity throughout their lifetime.

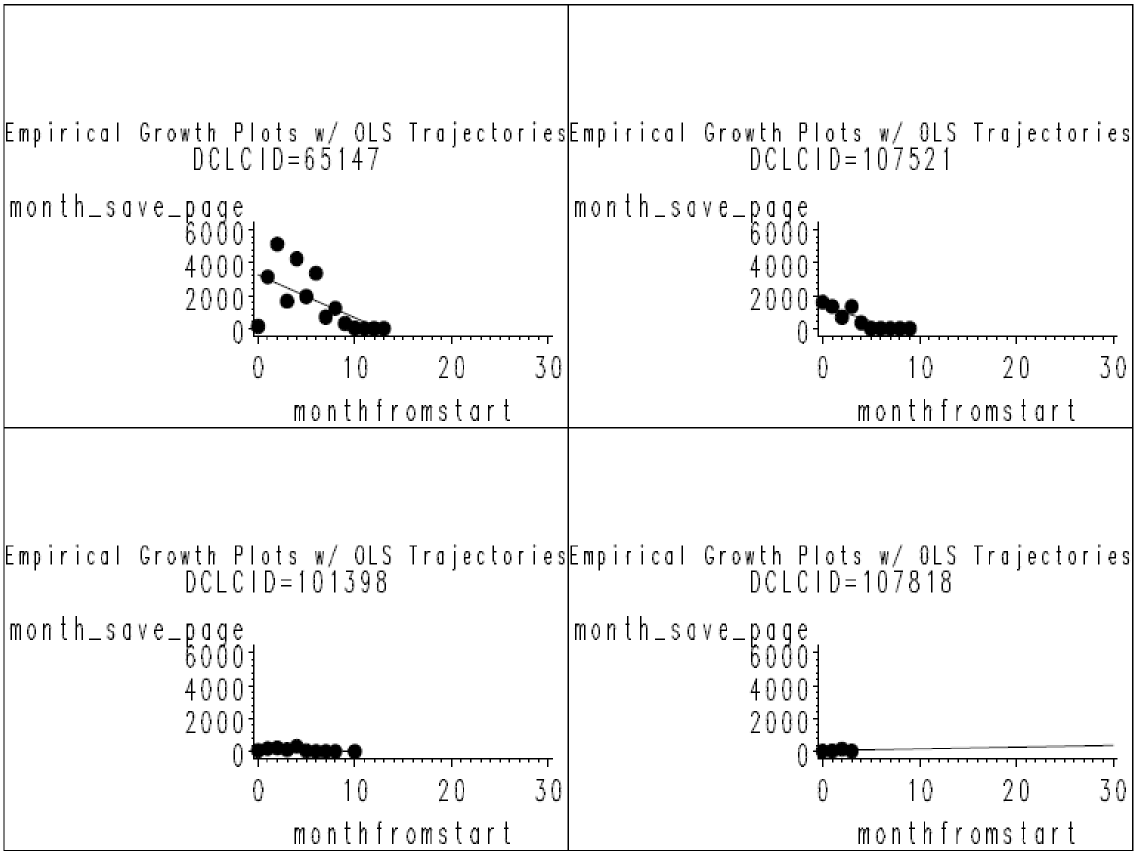

Figure 1: Empirical growth plots, measured in monthly page edits, for four U.S., K-12 wikis, with OLS trajectories.

We also created a scatterplot of days versus page edits on a single display, for the full sample, which we present here as Figure 2. This model is essentially an outlier analysis, since the density of points at low levels of wiki page edits cannot be ascertained from this figure. For instance, at any given day after day 1, the modal number of edits is 0; however, this density is not represented in the figure. Still, it is interesting to note that the two main “spikes” of activity occur in the very first days in the wiki lifecycle and then again around day 365, a year after wiki creation.

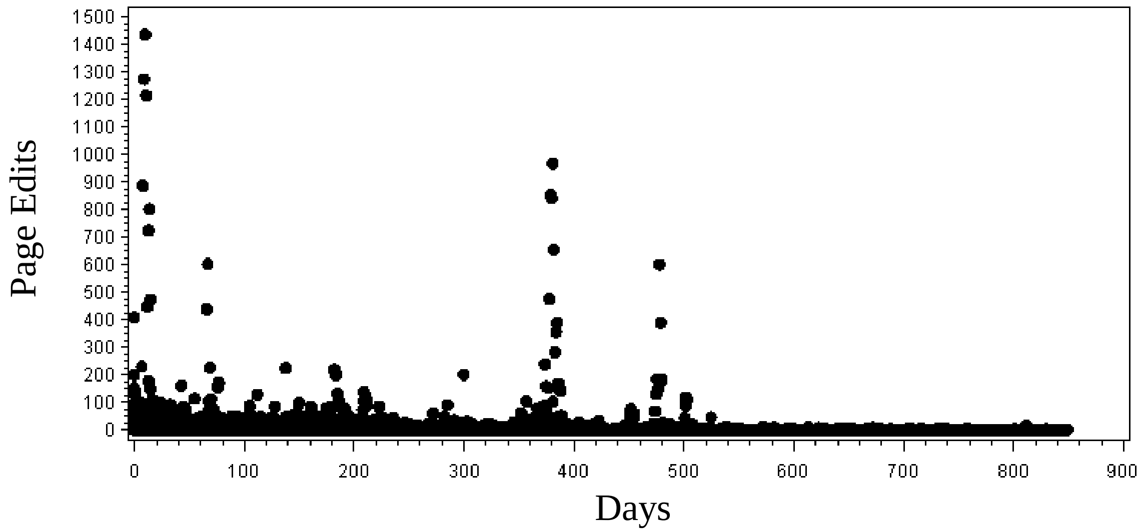

Figure 2: Scatterplot of page edits by day for U.S., K-12 Wikis (n=411).

After these empirical analyses, we summarized the longitudinal data using non-parametric loess smooth models. These smoothed regression models fit low-order polynomial functions to small subsets of the data and then connect these functions to create a smooth curve through the dataset. Since no studies of wiki communities existed to provide us any theoretical reasons for specifying a particular functional form of wiki developmental trajectories, the loess approach allowed us to model these trajectories without any parametric constraints.

One challenge of loess regression was that it required the specification of a smoothing parameter that determines the bandwidth of the local data subsets. If one chooses a bandwidth that is too narrow, then the fitted model features too much random variation; if one chooses a bandwidth too wide, then the model can smooth over important variation. Choosing an appropriate bandwidth is more art than science, since no robust methods exist for the optimal determination of the smoothing parameter. The goal is to choose a parameter that highlights the functional form of the curve and smoothes out the random variation.

In Figure 3, we present two loess curves with slightly different smoothing parameters. The panel on top has a narrower bandwidth than the panel on the bottom. Both panels generally have the same key features. First, on most days, the average number of edits across the U.S. K-12 wikis is very close to zero. On most days, most wikis experience no changes. Second, average wiki activity is higher very early in the wiki lifecycle and then a second spike occurs around day 365, after a year of wiki activity.

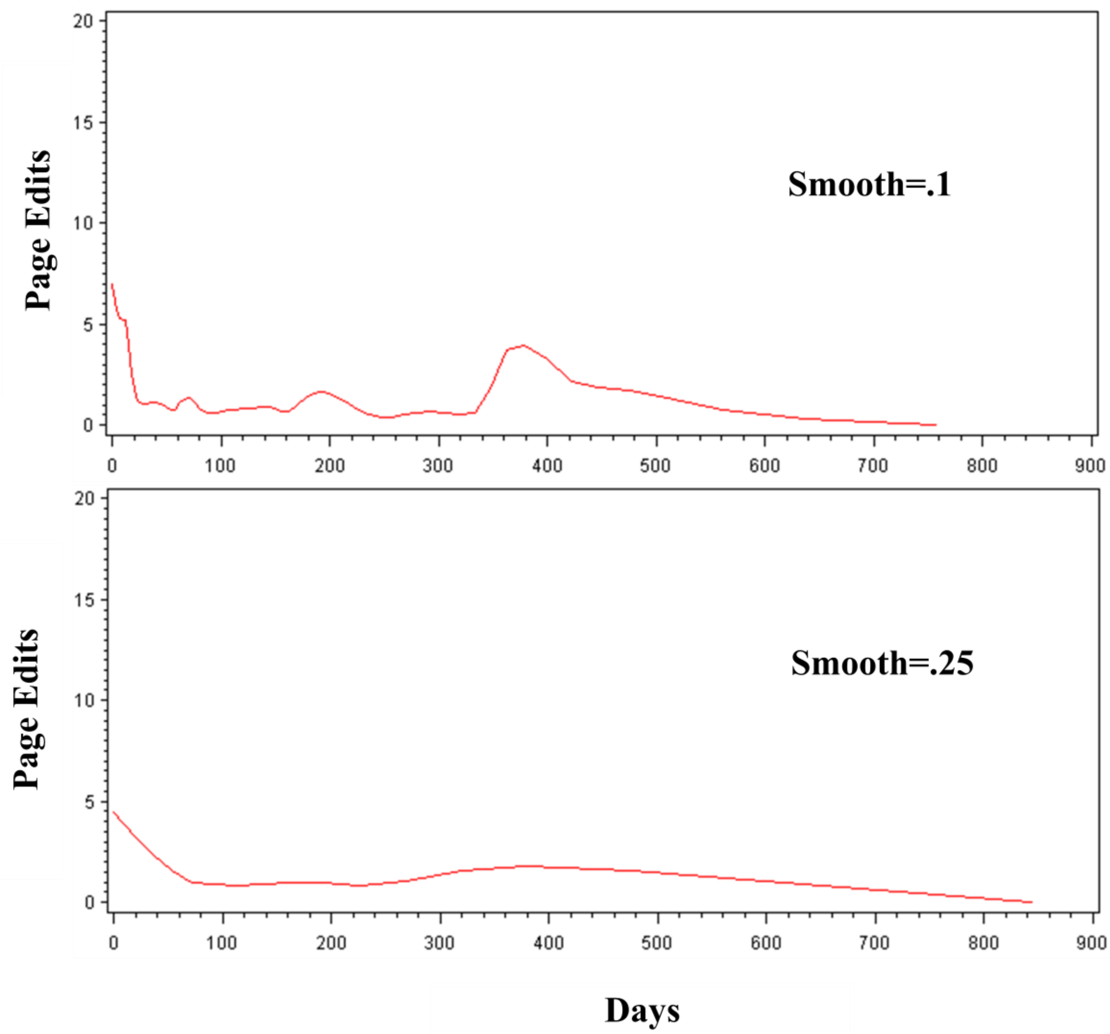

Figure 3: LOESS smooth curves of page edits by day for U.S., K-12 wikis, with smoothing parameters of .1 and .25 (n=411)

By inspecting these loess smooth plots, we drew three important conclusions about wiki developmental trajectories. First, on average, the bulk of wiki activity happens within the early days of wiki development. Second, there is an additional spike of activity after a year. Third, aside from these periods of high activity, on average, wiki development is negligible. On any given day, most wikis will experience few or no changes in page content, which suggests that they will not change in quality either. From these lessons, we determined that we would need to concentrate a considerable amount of our resources to measuring early wiki quality, and that we should also take at least one measurement after a year of activity to capture the spike that occurs around a wiki's one year anniversary, for those wikis that survive that long.

When in a Wiki's Lifecycle should we take Quality Measurements?

In order to determine the timing of wiki-quality measurements, we needed to assess typical wiki lifecycles. To do this, we needed to first define the features of a wiki lifecycle and definition rules for the "birth" and "death" of a wiki. With these decision rules in place, we used continuous-time survival analysis to model typical wiki survivor functions.

Measuring wiki lifetimes involves applying a biological metaphor, the lifecycle, to a socio-technical community. The birth of a wiki occurs at a distinct, measurable moment when a user generates a new subdomain on a wiki hosting network. Designating the moment of death of a wiki is more subjective, since wikis can always be returned to, changed and edited, even after years of inactivity. Nonetheless, we can identify precisely the last moment when a wiki was changed (through a page edit or new page creation), after waiting a sufficient time without further activity to ensure that the wiki is not merely dormant. Since the longest break in the U.S. academic year is the three-month summer holiday, we have adopted a 90 day period of inactivity as being sufficiently long to designate a wiki as "dead."

Other definitions of death are possible. One compelling alternative would be to choose the last time a wiki was viewed. We chose to use editorial changes because we value active engagement over the viewing of static information. We also could have chosen a different value for our 90-day observational window. In fact, we know that 13% of our 411 K-12 wikis have gaps between page edits that exceed 90 days, ranging from 92 to 754 days. During our survival analysis, we tested alternative models with a 120 day observational window and found that results did not differ substantially. Therefore, we decided against expanding the window further as we would have censored a large number of wikis very likely to be indefinitely inert in order to avoid labeling as dead a small number of wikis which may experience future changes.

With this established definition of a wiki lifetime, we could then model wiki lifetimes using survival analysis. We could not use simple univariate statistics to summarize lifetime or use wiki lifetime as an outcome in ordinary least-squares (OLS) regression analysis because of the problem of censoring. Not every wiki experiences the event of interest—the end of wiki activity—during our observational period. That is, some wikis have their final observed edit within our 90 day window, and as a result we did not know if these wikis are permanently inert or not. If we treated our lifetime measures as an outcome in OLS regression, our results would be biased since some wikis lived longer than our records indicate. Because of this issue of censoring, we use well-developed techniques from the literature on epidemiology known as “survival analysis” or “event history analysis.”

To conduct survival analysis, we reformatted the values of our measures in a project-level dataset, where every row in the dataset corresponded to one wiki. Event history analysis requires that we use a dichotomous measure of the event of interest to record the values of our outcome. Thus, we record *EVENT* as dichotomous variable coded as "1" when a wiki's final edit is at least 90 days before data collection and coded as "0" otherwise. Our measure of wiki lifetimes is *DAYS,* a continuous variable recording the number of days between creation date and last edit date. In our sample of U.S. K-12 wikis, *DAYS* ranged from 1 (meaning that the wiki's last change was within 24 hours of its creation) to 914.

As reported in our *Educational Researcher* paper (Reich, Murnane & Willett, in press), we used Kaplan-Meier methods to estimate "baseline" survivor functions for our 255 public school wikis. In Figure 4, we present the Kaplan-Meier estimated survivor function for our entire wiki sample (Singer & Willett, 2003). We display the time since wiki creation on the X-axis and estimated survival probabilities (the proportion of wikis that remain active beyond each particular time-point) on the Y-axis.

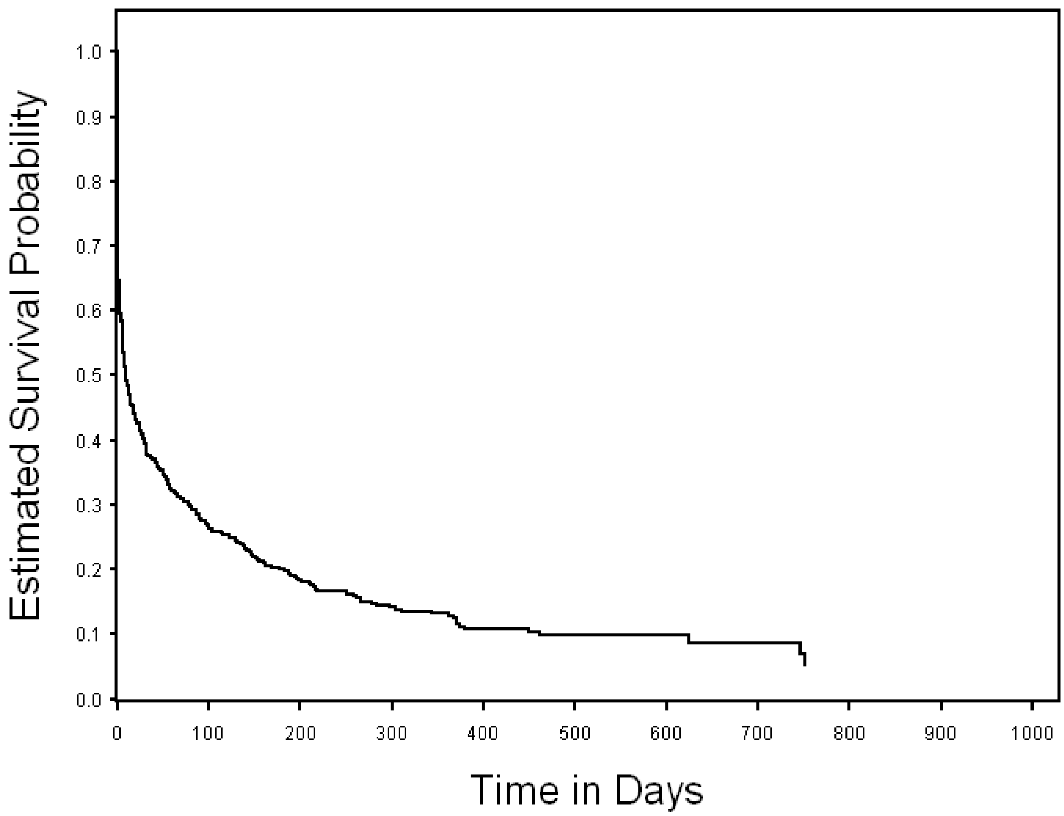

Figure 4: Estimated survivor function of U.S., K-12, public school wikis (n=255).

The steep initial drop in the estimated survivor function indicated that many wikis are terminated almost immediately after creation. For instance, the estimated median lifetime (the length of time beyond which 50% of the original wikis survive) of public school wikis was only 13 days, and only one quarter of wikis persisted beyond 151 days. These estimates suggested that most wikis that were used at all are used for short-term projects and assignments rather than serving as long-term course platforms or student portfolios.

From these findings, we established a useful summary statistic of wiki lifetimes, the median lifetime, that we could use as a referent for determining occasions of measurement. We also recognized that we would need to concentrate our measurements very early in the wiki creation process, since so many wikis failed after so few days.

The Protocol for Measuring Wiki Quality

Using the insights from our growth modeling and survival analysis, we established our protocols for collecting observations of wiki quality. In our earliest wiki quality coding, before we knew exactly how long it would take to code a wiki on average, we settled on four occasions of measurement: day 7, 14, 100, and 400. We chose days 7 and 14 in order to capture two occasions very early in the wiki lifecycle, given our knowledge that most wiki edits happen early in the wiki lifecycle and half of all public school wikis fail by day 14. It proved convenient that our media lifetime of 13 days was fairly close to a culturally-meaningful marker of time: 14 days or two weeks. While days 7 and 14 had methodological appeal, they are also easy to interpret at one and two weeks into a wiki's lifetime. We chose day 400 in order to capture the "bump" of activity that we found in wikis that survived at least a year. Finally, we added a third measure at day 100, to capture wiki quality at approximately the semester mark. We chose a date closer to the two week mark—rather than midway between 15 and 400 days—because we knew that wiki survival probabilities decrease rapidly over that period, and at day 100 we would still be measuring quality in approximately 25% of all wikis.

While the uneven spacing of these measurements may appear inelegant, there are good methodological reasons for choosing this spacing. The precision of ordinary least-squares regression estimates of rate of change in a growth model is a function of the precision of the measurements and the spacing and frequency of occasions of measurement (Singer & Willett, 2003). Additional measurements increase precision, and widely spacing additional measurements also increases precision. Therefore, the asymmetry in measurement actually increases the precision of our regression.

After some of our pilot studies in wiki coding, we determined that adding additional occasions of measurement would not be unduly expensive nor time consuming, especially measures taken after day 14 when 50% of wikis have ceased changing. We believed that additional measures would help us model complex wiki development more accurately. At the time of making this decision, we were concerned that sparse data at the higher values of time might cause difficulty when we tried to fit models of wiki-quality development with polynomial specifications of time. Therefore, we selected day 30 and day 60, days approximately twice and four times the median lifetime, as additional occasions of measurement.

In retrospect, choosing day 30 and day 60 as our additional occasions of measurement might not have been the best allocation of our resources. In developing our protocol, we assumed that we would use some kind of polynomial specification of time in modeling wiki quality. We also assumed that quality would develop throughout wiki's lifetimes and we needed to have sufficient data throughout the wiki lifetime to model potential complexities in these quality growth trajectories. These two assumptions proved to be incorrect. With six occasions of measurement, we determined that we had sufficient data to attempt complex, non-linear specifications of time. As became clear when analyzed our complete set of wiki quality measurements in our first sample, wiki quality is best modeled with a non-linear, logarithmic specification of time rather than a polynomial specification of time. Moreover, wiki quality changes primarily during the first two weeks of a wiki's lifetime, not throughout the entire wiki lifetime. We had predicted a concentration of activity early on from our wiki development trajectories, but the concentration of activity was even more striking than we hypothesized might occur. The most important days to measure wiki quality, therefore, are the earliest days of a wiki lifetime. Rather than adding day 30 and day 60, we would have perhaps been better off adding a measurement at day 1 to continue to increase our precision in modeling early wiki-quality development.

Thus, as we measured wiki quality in subsequent samples (for studies that we are currently working on), we evaluated quality at days 1, 7, 14, 30, 60, and 100. (We could not measure wiki quality at day 400 in these samples because the wikis had not persisted long enough at the time of our data collection.) In these subsequent studies, we hope to evaluate whether the additional measurement at day 1 improves the precision of our estimates.

We hope that other researchers can take away several lessons from this narrative of our development of protocols for the application of the WQI. First, we used questions concerning the timing and frequency of our measures to frame our decision-making about when to measure wiki quality. Second, in the absence of existing published research about wiki quality, we were able to use easily obtained data about wiki development to make reasonable assumptions about wiki-quality development. We learned from initial analyses that wikis typically survived for a short period of time and that most of their activity occurred early in the wiki lifecycle. Thus, we focused our resources on measuring the first days of wiki development, but we also took enough measures to track wiki quality growth over a full year. Finally, we used wiki-quality data from our first study to refine our protocols in subsequent studies (in particular, adding an additional occasion of measurement at day 1). In retrospect, it might have been wiser to conduct a small pilot study with a random subsample of our 255 public school wikis, and completely analyze the quality measures before pressing ahead with a complete study of our first sample. We did not do so because by the time we had trained our research assistants, we needed to keep them working steadily throughout the year, so taking a break from measuring while we conducted in depth analyses would not have been feasible. It might have been wiser to plan for a small but complete pilot from the beginning, and manage our hiring and staffing accordingly. Overall, however, our strategies for designing the WQI protocols allowed us to address the research questions of our larger study.

Part V: Adaptation Guidelines for Educators

Since we began presenting our research on evaluating wiki quality, educators have asked us to share our Wiki Quality Instrument (WQI) and our thoughts on assessing wiki quality for classroom purposes. We designed the WQI for the large scale evaluation of wikis used for diverse purposes in diverse educational settings, so in its current form the WQI is not a suitable tool for assessing individual students in particular classrooms. We do believe, however, that the WQI has some application for assessing technology use in schools and school districts. We also believe that the categories of the WQI could serve as a useful conceptual framework for teachers designing rubrics for evaluating student performances in online learning environments. In this section, we describe several ideas that we have for adapting the WQI for educational settings. First, we provide a bit of background on the WQI as a way of introducing some of its strengths and limitations. Second, we suggest possible ways of using the WQI to assess technology use at the district and school levels. Then, we present findings from our research about how teachers' assess wiki quality as an introduction to our suggestions for using the WQI as a conceptual framework for developing wiki rubrics.

Before beginning, it is important to note that our early research agenda was designed to assess wiki usage in schools, not to be able to translate those findings directly into actionable advice. What follows therefore is advice which is not the result of a set of research studies designed to tell educators how to evaluate wikis; rather it is advice that emerges from people who have spent years thinking about wikis in schools and are extrapolating from general experience rather than specific research findings. (By way of analogy, I recently went to my orthopedic surgeon for a check-up several years after knee surgery to repair a traumatic injury to my knee cartilage. He suggested that I was now well enough to participate in sports, but I should probably not pick up marathon running. He emphasized that his advice was not based on specific scientific research; there are not randomized controlled trials evaluating the impact of marathon running on victims of traumatic knee cartilage injuries. Rather the foundation for his advice came from spending a lot of time looking at and thinking about knees. The advice herein is in a similar vein, based on our personal insights rather than findings from investigations of specific research questions.)

Background on the WQI

The WQI was designed to evaluate an extraordinarily diverse set of learning environments. We examine wikis used in pre-K classrooms and senior electives, in AP classes and ESL classes, in every subject area imaginable, all across the country and in U.S. schools overseas. Across these different sectors, wikis are used for all kinds of purposes: as teacher Web sites, electronic worksheets, individual presentations, collaborative workspaces, and portfolios. Moreover, we can only see these learning environments from one perspective: the wiki itself. We know that every wiki we analyze is nested within a larger ecology of learning, an ecology which we cannot examine in its entirety.

To consistently measure dimensions of activity across these diverse spaces, our instrument takes very coarse measures. We attempt to identify the presence or absence of 24 behaviors that appear across different types of wiki learning environments that provide opportunities for students to develop 21st century skills. During the course of our pilot testing, we tried several other approaches. We attempted to consistently measure the frequency of a behavior, but we were stymied by the diverse size of wikis. If something happens 20 times on one page, is that more frequent than 2 times each on 8 different pages? We also attempted to qualitatively assess certain actions; for instance by trying to measure whether typeface formatting was used “decoratively” or “substantively;” We found that our research assistants could not consistently agree on what constituted substantive use of formatting. As a result, the WQI does not necessarily evaluate whether or not high quality student work is happening. Since we cannot reasonably compare the quality of copyediting done by third graders with the copyediting done by twelfth graders, we do not attempt to evaluate the quality of copyediting at all. Rather we assess whether or not there is any copyediting to be found. The WQI instrument is best understood as a tool to identify whether or not certain preconditions for 21st century learning are present on a wiki.

For districts and schools that are in the early phases of evaluating technology initiatives, identifying these kinds of preconditions might prove to be a very useful exercise. The WQI offers a cost-effective, relatively simple method for evaluating whether students are using wikis in ways that could potentially foster the development of 21st century skills. For classroom teachers who have specific learning goals in mind for students, the broad indicators of the WQI are less useful. They may, however, be useful for teachers who are using wikis without clear goals, who need to develop specific benchmarks of quality. In the sections that follow, I suggest how school leaders and classroom teachers might apply or adopt the WQI for different circumstances.

Before beginning, it is important to note that WQI in its present form would only be useful in evaluating student activity with wikis. While other media are similar, such as blogs or the Google Docs package, the affordances of these tools does not align exactly with the affordances of wikis. For instance, it is not really possible to “copyedit” an individually-authored response to a blog post, and there are functions that can be performed with Google Docs and blogs that cannot be accomplished with wikis. At this time the WQI is well suited only to evaluate the use of wikis. In Part 6: Adaptation Guidelines for Researchers, we discuss suggestions for using the WQI as a template for devising other quality instruments for online learning environments. Here, however, we restrict ourselves to only examining the use of wikis in schools and districts.

Using the WQI to assess wiki usage in schools and districts

A school or district administrator charged with evaluating the efficacy of school or district-wide technology initiatives faces many of the same challenges that faced our research team as we sought to evaluate wiki usage at a national scale. Across a district, wikis can be used in many grade levels, in many subject areas and for many purposes. The WQI is one tool that could be used or adapted to assess the degree to which wikis in a school or district provide opportunities for 21st century skill development.

Ideally, a school or district-wide assessment of wiki usage should evaluate the degree to which wikis support a school or district’s learning goals related to technology. Unfortunately, in our experience, many American schools and districts do not have organized goals for learning technologies. Instead they support radical teacher autonomy and let 1,000 flowers bloom.

To highlight this common omission in district technology planning, it is helpful to look at school systems that do have clear visions for how technology should support student learning. For instance, in 2008 the Ministry of Education in Singapore launched the Third Masterplan for ICT in Education (http://ictconnection.edumall.sg/), which directed educators across Singapore to focus on using technology to support only two learning goals: to develop student capacities for self-directed learning and collaborative learning. There are, of course, other learning goals that technology can support. However, by choosing to focus on just two, Singapore has the opportunity to develop a coherent approach to teacher training, student evaluation and technology investment across the nation's schools.

American school districts, in our experience, almost never have this clarity of focus around technology investment. The norm is probably to have no clear goals for how those investments should improve student learning outcomes. Often times, technology is viewed as an end in itself—where teachers use more technology so that students learn more about technology—rather than as a suite of tools to help students develop mastery of more important learning goals. In some districts, there are vaguely articulated goals for technology use, sometime related to ideas around 21st century thinking, but these goals lack shared definitions of key terms and shared outcome measures.

For our purposes here, let us imagine how three different types of districts might use the WQI. The first district has very clearly demarked goals for student outcomes from technology investments, the second district has a vague notion that technology should support 21st century learning but no institutional consensus or conception of what that means, and the third district has defined goals for their technology investments.

For districts with clearly demarked goals for student outcomes from technology investments, it might make sense to use only those parts of the WQI that correlate with district goals. For instance, to continue with the example of Singapore, educators in that country might use the Complex Communication subdomain to capture collaborative learning opportunities with wikis. Some of the questions on the Expert Thinking subscale might capture certain elements of what Singaporeans mean by “self-directed learning,” but those items which do not correlate could be discarded. If Singaporeans are not interested in the development of New Media Literacy, that entire scale could be ignored. The advantage of discarding unnecessary parts of the WQI is twofold. First, a shortened instrument will require less time to apply. Second, analyzing elements unrelated to the core learning goals of a district can lead to distractions and tangential analyses. With the limited time available for district staff to analyze student work, it makes sense to focus carefully on school-wide and district-wide goals.

For districts with fuzzy goals or without clear learning goals around technology use, using the entire WQI might provide a kind of “baseline assessment” of the ways in which wikis are used throughout a school or district. This kind of data can be useful in jumpstarting conversations around learning goals for technology interventions and investments. For instance, if a school found that its use of wikis paralleled national trends, with most wikis serving as individual content delivery devices with minimal collaboration, that could open conversations about the kinds of support teachers need to conduct more innovative work. If wikis in a school scored low on the WQI across the board, that could also stimulate conversation about why teachers were using wikis and what help they need to use them successfully. If a school's wikis scored high in a particular dimension, then it could be sensible to focus on improving outcomes within that domain. If a school's wikis score highly in several domains but not in another, then it could be sensible to focus on the weak area. The best pathway for improvement would depend upon a wide variety of factors: how much capacity the school has to take on an instructional initiative, how important wikis are to teaching practice, the percentage of faculty members using wikis, the support that can be engendered for this kind of work among key stakeholder groups, and so forth. An analysis of wiki quality can provide some baseline data for school and district administrators to evaluate whether or not wikis are providing opportunities for students to develop 21st century skills. That baseline data can then be used as a shared text to ground a conversation about next steps for the district.

For districts that are exclusively focused on raising scores on standardized tests of reading comprehension and mathematics, the WQI is unlikely to be of much help. There are not compelling reasons to believe that opportunities to develop expert thinking, complex communication, and new media literacy will be correlated with gains of scores on state tests.

In terms of the practicalities of using the WQI in a school or district self-study, the procedures that we outline above could be simplified and used within a district. The first step would be collecting a district-wide sample of wikis. If the district has a district-wide wiki provider, like a contract with PBworks or Wikispaces, then obtaining a sample of wikis is simply a matter of randomly sampling from the list of all wikis created on the wiki service. If teachers are individually creating wikis with their own accounts, then the best strategy would be to survey all teachers in the district and ask them to share their wiki URLs, and then sample from this list. (An alternative approach would be to randomly sample a set of teachers, and personally request wiki URLs from them. This might be more effective if teachers are unlikely to complete district surveys). How big should the sample be? In all likelihood the optimal number is higher than the number that a district can afford to evaluate, so districts should sample as many as they can afford the time to examine. If very few wikis are used in the district, a team might be able to examine all of them. But if a team cannot examine every wiki, a randomly drawn sample will work fine. Bigger samples are better, but small samples are better than doing no assessment at all.

With a sample of wikis identified, the next step would be to identify which parts of the WQI to use. Use only the quality subdomains that align with district goals. If the district has no goals, consider using the whole instrument as a mechanism for developing some baseline data about technology use. Most districts should consider collecting demographic data in terms of which subject areas and grades levels are supported by the wikis, so that the district can analyze wiki usage in different parts of the organization.

The protocols described in Part II of this document lay out a sophisticated series of strategies for using multiple raters to evaluate wiki learning environments and for resolving disagreements. While our protocols call for wikis to be independently rated by two coders, districts may not be able to devote the resources to fully using our protocols. Having only one rater evaluate each wiki is not perfect, but it is better to have one rating per wiki than not doing any assessment at all. Educators should not let the perfect be the enemy of the good.

Schools and districts interested in conducting these kinds of study are welcome to get in touch with the first author at justin@edtechteacher.org for further advice and counsel on this process. In most cases he can offer some guidance over the phone, and he may be available for consulting or partnering on a larger research study.

Adapting the Wiki Quality Instrument for Classroom Use

In this section we discuss how the WQI could be adapted by classroom teachers for evaluating student activity and learning with wikis. Before directly considering adapting the WQI for K-12 settings, it is worth providing some context about how classroom teachers typically evaluate wiki quality. In our research, we spent a year conducting qualitative research in U.S. schools trying to answer the question “How do K-12 teachers define and assess wiki quality?” We found that teachers often could articulate clear learning goals for their wikis: they hoped that wikis would help students develop collaboration skills, facilitate course logistics, build students’ technological fluency, and allow students to deepen and demonstrate understanding of course materials.

That said, when we asked teachers and students how high quality work on wikis was evaluated, we got a different set of answers. Teachers told us that they primarily evaluate students on whether they contributed to the wiki at required intervals, whether they included factually correct content, and whether they met project guidelines. In short, rather than evaluating students on whether they demonstrated mastery of the stated learning goals, teachers evaluated students on whether or not they followed the directions of the assignment.

The items of the WQI can help teachers think about new ways to assess 21st century skills in the domains of expert thinking, complex communication, and new media literacy. The items will probably not be useful to teachers by themselves. The Wiki Quality Instrument was designed to evaluate wikis across a diverse population of wiki used in many subjects, grades, and schools. As a result, our measures are very coarse, compared to the fine grained measures that teachers should use to evaluate their students. For instance, we use the WQI to assess whether or not copyediting occurs on a wiki. We do not assess whether students do that copyediting thoroughly or effectively, we simply note if the activity exists. Teachers, however, should be interested in whether a student copyeditor assesses an entire piece thoroughly, finds a high percentage of errors, corrects them accurately, and so forth.

Thus, the items of the WQI are probably better suited to helping teachers brainstorm rubric categories than for generating benchmarks in specific categories. For instance, imagine a teacher who is hoping to have seventh grade students in an Earth Science class use wikis to develop collaboration skills. One way to begin developing a rubric for her wiki project would be to look at the Complex Communication items in the WQI. She could then decide which categories would be relevant to her rubric. For instance, she might decide that students have no need to use the wiki for scheduling, so that should be left out of the rubric. She might decide that commenting is essential to the project, and then develop criteria for four benchmarks: Exceeds Standards, Meets Standards, Approaches Standards, Just Beginning. The Exceeding Standards benchmark might be defined as "Providing specific, constructive feedback on multiple pages for all collaborators," and a Just Beginning might be defined as "Adds comments to at least one collaborators page, but comments are vague general praise or inappropriate criticism." Certain aspects of the standards might be tied even more specifically to the wiki project goals or to the Earth Science content.

This teacher also might decide to create one rubric category drawing on several items from the WQI. For instance, she might have one category called "Collaboration", where the Just Beginning benchmark describes concatenation, the Approaches Standard benchmark describes copyediting, the Meets Standards benchmark describes co-construction, and the exceeds standards benchmark describes specific performance criteria for high quality co-construction in the context of the wiki project goals.

The WQI has potential as a source of inspiration for classroom teachers as they develop rubrics for work on wikis, but by itself it will not be sufficient. Wiki-using educators need to think carefully about their learning goals, and then develop strategies for communicating those goals to students. The WQI offers a framework for breaking down 21st century skills into some of their constituent parts. Those parts may be useful in inspiring teachers to consider how to develop rubrics for wiki work that evaluate the degree to which students demonstrate mastery of the learning goals associated with a wiki project.

Part VI: Adaptation Guidelines for Researchers

We believe that the online data warehouses of Web 2.0 learning environments provide an extraordinary opportunity for researchers to study the processes of learning in real-time and at scale. Large scale content analysis using instruments like the WQI represents one method of mining this new vein of data. We hope that other researchers will consider conducting studies similar to ours and help map the evolving landscape of social media use in classrooms.

Applying the Current Version of the WQI in Alternate Settings

Certain kinds of research could be conducted using the WQI without any modification. For instance, researchers could look at the use of PBworks wikis in K-12 context in other countries, especially if those education systems have strong similarities to the U.S. system. It also may be possible to use the WQI to assess wikis used in higher education contexts, although it may be that certain kinds of measurable wiki behaviors appear in higher education contexts that we did not find in our analysis of K-12 settings. If this is true, then the new items would need to be developed to assess those behaviors.

Modifying the WQI to Study other Wiki Providers

Most additional studies of the use of social media and wikis in K-12 settings, however, will probably require modifying or adapting the WQI. One promising research study would be to compare behaviors on Wikispaces with patterns of behavior on PBworks. Our hypothesis is that the two platforms would be used in very similar ways, but that hypothesis deserves to be tested. Since the Wikispaces platform has slightly different affordances than the PBworks platform, certain items might need to be adjusted. In particular, Wikispaces has no special place for comments on a content page, but it does have a threaded discussion board associated with every content page. Thus, the criteria for commenting and discussion might need to be altered.

To make these modifications, it would first make sense to search the literature and examine whether other scholars have analyzed these elements of Wikispaces. If so, the research methods of other scholars might provide guidance as to how best to analyze opportunities to develop 21st century skills on the Wikispaces platform. After this step, we recommend analyzing a random sample of perhaps 200 wikis from the population of interest and conducting an open-ended round of coding where researchers briefly describe the ways in which teachers and students using the discussion pages and content pages for commenting and discussion. From these descriptions, researchers can craft new items for the WQI or determine that the original version of the WQI adequately describes measurable behaviors on Wikispaces wikis. This same process would work for wiki platforms other than Wikispaces, such as a proprietary wiki solution within a university or Content Management System(CMS) platform or an installation of MediaWiki in a district or university.

Modifying the WQI to study Wiki Usage in Particular Curricular Domains

The further afield researchers go from the original purposes of the WQI, the more the instrument will need to be adapted. One logical direction for this line of research is to look at wikis used in specific contexts, such as in particular academic subject (Earth Science), a particular grade (7th grade), a particular set of schools (Iowa public schools), or some combination of these factors (7th Grade Earth Science classrooms in Iowa.) The WQI, without modification, could be used in such a context to coarsely measure opportunities for students to develop 21st century skills. However, in more specific settings, we believe that it is considerably more feasible to develop more fine grained measures of the quality of these 21st century skill learning opportunities. For instance, we found it extremely challenging to develop a taxonomy of behaviors associated with the development of expert thinking that might apply across grade levels and across subject areas. Within Earth Science, however, researchers might be able to identify a series of discursive moves on Earth Science wikis that represent students applying scientific thinking skills: posing questions, generating hypotheses, presenting evidence, testing hypotheses with evidence, designing experiments, and so forth. Researchers could also identify when students work with course content that aligns with state standards or extends beyond those standards.

It might be that these kinds of discursive moves could be added as additional dichotomous items to the WQI (Do students generate hypotheses concerning Earth Science phenomena?) However, it might also be possible to go even further and measure the actual quality of these discursive moves. Researchers could develop a rubric of criteria distinguishing high-quality and low-quality hypothesis generation as established by national standards, state standards, or other research efforts. Our various efforts at scalar measures of quality failed because the diversity of our sample was too great—how would one compare copyediting quality of 11th graders working on poetry with 3rd graders working on local history? Within a more specific context however, it might be possible to assess not just whether 7th graders in Iowa use their Earth Science wikis to generate hypotheses, but to evaluate the quality of their efforts. Similarly, researchers could evaluate not just whether 7th graders in Iowa use academic content about soil conservation, but whether they demonstrate mastery of Iowa 7th Grade Earth Science Standard 3, Interval Benchmark 2d: “Knows conservation methods that lessen the effects of soil erosion.”

If we were to attempt such as study, we would generally follow the process that we describe in the section on Part III: Developing the WQI, with a focus on these particular wikis in Iowa. We would begin with three overlapping methods for considering 7th grade Earth Science wiki quality. First, we would conduct observational studies and interview research with wiki using Earth Science teachers and students to evaluate how they use wikis, the role that wikis play in their classrooms, and the ways in which teachers and students define and assess high quality work in wikis. Second, we would conduct a literature review to identify existing measures of quality in online learning environments for science or science learning broadly. Third, we would assess our own theoretical frameworks for high quality work in these wikis, so as to be cognizant of our own beliefs and biases.

From these three methods, we would attempt to identify the domains that we believed would be worthy of analysis. It might prove that participation, information consumption, expert thinking, complex communication, and new media literacy remain germane categories for these Earth Science wikis. It seems very likely that we might need to refine our categories, such as making scientific thinking an explicit dimension of expert thinking or perhaps its own subdomain. It might be, however, that teachers and students discuss very different criteria for evaluating wiki quality and therefore entirely different domains of analysis are warranted.

With a better understanding of how wikis are used in their contexts in the instructional core, it might then make sense to examine a large set of Earth Science wikis to get a better sense of how these wikis are used. One of our first steps in developing the WQI was conducting a round of open coding on our sample of 1,799 wikis where we had researchers briefly describe how the wiki was being used. We conducted a second round of open coding on all 411 U.S. K-12 wikis to generate descriptions of behaviors that coders thought could promote the development of 21st century skills. From these qualitative descriptions, we began to develop decision rules for items in the WQI. For instance, the descriptions of various ways that students collaborated and communicated on wikis informed our development of the seven complex communication items on the WQI. A similar protocol might be used in a specific domain like Earth Science, where researchers examine a set of relevant wikis to begin to identify common discursive moves on Earth Science wikis that can be measured systematically. With a set of preliminary items developed from these efforts, researchers could then pilot test the items and determine their validity and reliability, in a process of developing a refined 7th Grade Iowa Earth Science Wiki Quality Instrument.

Adapting the WQI for the Study of Other Social Media Platforms in K-12 Settings

All of these suggestions thus far have involved adapting the WQI to be used in evaluating wikis in different settings. We also hope that researchers consider conducting similar kinds of large-scale content analysis on other kinds of online learning platforms.

Public discussion boards, blogs, niche social networks like Ning, media sharing and commenting platforms like Voicethread, Google Docs, and other emerging social media platforms are all promising sites for exploring student learning in online venues. Certain items from the WQI might be adapted relatively easily to these new contexts; for instance blogs and wikis offer similar affordances for communicating with new media. Google Docs and wikis offer a similar breadth of possibility for collaboration. Platforms that are different from wikis, such as the media sharing and commenting site Voicethread, might require a very different analytical approach from the one that we took in analyzing wikis.

In addition to developing new instruments for evaluating quality in other online learning environments, researchers will also need to develop protocols for evaluating quality in those environments. To develop our protocols for measuring wiki quality, we needed to address the questions of "how often should we measure quality?" and "when should we measure quality?" Part III: Developing WQI protocols explains our processes for determining these protocols in wikis. Briefly, the question of how often to measure quality needs to be based on the complexity of typical quality development trajectories. If these trajectories are linear, then three data points are sufficient to capture the trend. The more complicated the trajectory, the more data points are needed to accurately represent the trajectory. We used wiki page edits to provide a basic model of wiki development trajectories. To answer the question of when should we measure quality, we used survival analysis to determine the range and median lifetime of wiki lifecycles. Findings from these analyses informed our decisions about when to measure wiki quality. While it might be reasonable to adopt our protocols for measuring other wiki hosting platforms, for researchers adapting the WQI for other platforms, it would be necessary to develop new protocols for measuring quality in these new domains.

Conclusion

Whether or not the specific items from the WQI can be adapted to other platforms, we hope that our general methodological approach can be adapted for use in the study of other public online learning environments. The key features of our work are that we:

1) Access a population of online learning environments with extensive real-time data about individual behaviors
2) Randomly sampled from those environments to study a representative sample of the population
3) Devise protocols for the frequency of quality measures by assessing typical developmental trajectories for the online learning environments.
4) Devise measures of quality. These measures emerge both from what teachers and students believe represents high-quality work and from what the literature says about elements of high-quality learning.
5) Treat quality as time-varying, and measure quality at multiple time points throughout the lifecycle of the online learning environment. Treat quality as time-varying rather than a static feature of the online learning environment.
6) Evaluate how covariates associated with the learning environment (teacher attitudes, school resources, student characteristics, etc.) affect the initial position and rate of change of the quality trajectories.

Our approach has yielded important insights about who uses wikis, how they are used, the kinds of learning opportunities that students have with wikis, and how these learning opportunities are distributed across schools serving different populations. We believe that similar approaches applied to other online learning environments could prove to be similarly productive, and we are available to consult with other researchers conducting these kinds of studies.

References

Charmaz, K. (2006). *Constructing grounded theory : A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage Publications.

Chiu, M. M. (2008). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. *Journal of the Learning Sciences, 17*(3), 415.

Cortez, C., Nussbaum, M., Woywood, G., & Aravena, R. (2009). Learning to collaborate by collaborating: A face-to-face collaborative activity for measuring and learning basics about teamwork. *Journal of Computer Assisted Learning, 25*(2), 126-142. doi:10.1111/j.1365-2729.2008.00298.x

Dede, C. (2010). Comparing frameworks for 21st century skills. In J. A. Bellanca, & R. S. Brandt (Eds.), *21st century skills : Rethinking how students learn* (pp. 51-76). Bloomington, IN: Solution Tree Press.

Gardner, H. (2006). *Five minds for the future; 5 minds for the future*. Boston, MA: Harvard Business School Press.

Haste, H. (2008). Constructing competence: Discourse, identity and culture. In I. Plath, I. Graudenz & H. Breit (Eds.), *Kultur — handlung — demokratie* (pp. 109-134) VS Verlag für Sozialwissenschaften.

Jenkins, H. (2006). *Confronting the challenges of participatory culture: Media education for the 21st century.* Chicago, IL.: MacArthur Foundation. Retrieved from www.digitallearning.macfound.org

Jenkins, H. (2009). *Confronting the challenges of participatory culture : Media education for the 21st century*. Cambridge, MA: The MIT Press.

Jeong, A. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussions. *American Journal of Distance Education, 17*(1), 25-43.

Kozma, R. B. (2003). Technology and classroom practices: An international study. *Journal of Research on Technology in Education, 36*(1), 1-14.

Lawrence, N. (2009). *The effects of using a wiki as a tool to complete an academic task on collaborative skills of disadvantaged, urban youth: A pilot study.* (M.S., University of Kansas).

Levy, F., & Murnane, R. J. (2004). *The new division of labor : How computers are creating the next job market*. Princeton, N.J.: Princeton University Press.

Moskaliuk, J., Kimmerle, J., & Cress, U. (2009). Wiki-supported learning and knowledge building: Effects of incongruity between knowledge and information. *Journal of Computer Assisted Learning, 25*(6), 549-561. doi:10.1111/j.1365-2729.2009.00331.x

Oriogun, P., Ravenscroft, A., & Cook, J. (2005). Validating an approach to examining cognitive engagement within online groups. *American Journal of Distance Education, 19*(4), 197-214.

Partnership for 21st Century Skills. (2007). *The intellectual and policy foundations of the 21st century skills framework.* Retrieved from http://www.21stcenturyskills.org/route21/images/stories/epapers/skills_foundations_final.pdf

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis : Modeling change and event occurrence*. Oxford ; New York: Oxford University Press.

Trentin, G. (2009). Using a wiki to evaluate individual contribution to a collaborative learning project. *Journal of Computer Assisted Learning, 25*(1), 43-55. doi:10.1111/j.1365-2729.2008.00276.x

Trilling, B., Fadel, C., & Partnership for 21st Century Skills. (2009). *21st century skills : Learning for life in our times; twenty first century skills*. San Francisco: Jossey-Bass.

Wagner, T. (2008). *The global achievement gap : Why even our best schools don't teach the new survival skills our children need--and what we can do about it*. New York: Basic Books.

VITA

Blair Justin Fire Reich

1996-1999	University of Virginia Charlottesville, VA	B.A. June 1999
1999-2000	University of Virginia Charlottesville, VA	M.A. June 2000
2000-2002	Instructor, SOLO Conway, NH	
2002-2003	Director of Operations, Shackleton Schools Ashby, MA	
2003-2006	Teacher, Noble and Greenough School Dedham, MA	
2006-Present	Co-director, EdTechTeacher Chestnut Hill, MA	
2007-2012	Doctoral Candidate, Graduate School of Education Harvard University	
2008-Present	Project Manager Distributed Collaborative Learning Communities Project Graduate School of Education Harvard University	