Mind the Skills Gap:
The Role of Internet Know-How and Gender in Contributions to Wikipedia

Eszter Hargittai and Aaron Shaw*
eszter@northwestern.edu & aaronshaw@northwestern.edu
Northwestern University

Abstract

Despite the egalitarian rhetoric surrounding online cultural production, profound gender inequalities remain in who contributes to one of the most visited Web sites worldwide, Wikipedia. In analyzing this persistent disparity, previous research has focused on aspects of current contributors and the existing Wikipedia community. We draw on unique panel survey data of young adults with information about both Wikipedia contributors and non-contributors. We examine the role of people's background attributes and Internet skills in participation on the site. We find that the gender gap in editing is exacerbated by a similarly significant Internet skills gap. Our results show that the most likely contributors are high-skilled males and that among low-skilled Internet users, no gender gap in Wikipedia contributions exists. Our findings suggest that efforts to understand the gender gap must also take Internet skills into account.

Keywords
Digital inequality, Gender, Internet skills, Participation, Peer production, Wikipedia

Introduction

The growth of online communities and social media has inspired enthusiasm about new opportunities for people to express themselves and participate in the production of freely-available cultural goods (e.g., Benkler, 2006; Jenkins, 2006). Indeed, millions of people worldwide have contributed their time and effort towards the creation of public resources such as Wikipedia, Free/Libre Open Source Software, social question-and-answer sites like Yahoo! Answers, StackOverflow, and more. Yet, research has consistently found that women are less likely to contribute to many such endeavors than men despite being online at similar rates (Zickuhr and Smith, 2012), with especially robust results found among Wikipedia editors (e.g., Antin, Yee, Cheshire, & Nov, 2011; Benkler, 2002; Cohen, 2011; Collier & Bear, 2012; Glott, Ghosh, & Schmidt, 2010; Hill & Shaw, 2013). Given that large numbers – over 87 million unique monthly visitors in 2013 – of Internet users consult Wikipedia regularly, its content reaches a vast audience (ComScore, 2013). Accordingly, gender disparities in who authors the material on the site are increasingly a matter of public concern as the content disproportionately reflects the interests and perspectives of its most active contributors who are mainly men. This "gender gap" in contributions has persisted despite public outcry (Eckert & Steiner, 2013; Filipacchi, 2013; Gleick, 2013). Campaigns by the Wikimedia Foundation intended to encourage the retention of new female editors (e.g., Morgan, Bouterse, Walls, & Stierch, 2013) have not resulted in vast changes.

The causes and mechanisms that sustain the Wikipedia gender gap – as with other forms of gender inequity in society – are likely complex. However, most previous research on the Wikipedia gender gap has looked for answers *within* the community of Wikipedia editors, analyzing the community culture, the dynamics of contribution, and testing design solutions intended to mitigate the problem among current participants (e.g., Antin et al., 2011; Halfaker, Kittur, & Riedl, 2011; Morgan et al., 2013). We contend that these approaches are an important step, but are necessarily

limited because they only consider existing Wikipedia contributors when seeking to understand or resolve participation inequalities. Some of the strongest mechanisms of exclusion likely concern barriers that prevent initial contribution, leading to differentiation in who becomes an editor in the first place (Collier & Bear, 2012). Accordingly, by limiting studies to those who have at one point or another already contributed, research misses an important part of the puzzle: what factors explain who selects into contributing to Wikipedia in the first place?

Differential rates of cultural participation are an important line of sociological inquiry as they have been linked to social inequality both in terms of who participates and the resulting benefits of participation (e.g., Alderson, Junisbai, & Heacock, 2007; DiMaggio & Useem, 1978; DiMaggio & Ostrower, 1990). Traditionally, participation has mainly consisted of audience membership, e.g., museum or concert visits, since contributing content to public discourse, while possible (Perrin & Vaisey, 2008) had much more considerable barriers to entry. In the digital age, new modes of content production are available with lower barriers to entry and larger potential audiences than was previously within reach (Hargittai, 2000), but inequalities have nonetheless persisted in who contributes written, audio and visual content to the Web (Blank, 2013; Correa, 2010; Hargittai & Walejko, 2008; Schradie, 2011).

Among the many potential barriers to initial participation, we focus our analysis on the role of Internet experiences and skills. In doing so, we build on an extensive body of research considering the role of Internet experiences and skills in conjunction with various demographic and socioeconomic factors, in shaping what people do online (Hargittai & Hsieh, 2013; Howard, Rainie, & Jones, 2001; Schradie, 2011). This previous work has shown that comparable access to information resources and online tools does not necessarily lead to similar online behavior (Zillien & Hargittai, 2009). A persistent "skills divide" separates those who engage in many online behaviors from those who do not (Hargittai & Hinnant, 2008). Of particular relevance to the case of the

Wikipedia gender gap, research on differentiated Internet experiences and skills has shown that socioeconomic inequalities and demographic differences tend to translate into differentiated patterns of Internet use across cultural, civic, and relational spheres of activity (see Hargittai & Hsieh, 2013 for a review of related literature). We have every reason to believe that this might also be the case with Wikipedia and the gender gap.

Based on earlier work highlighting the relationship between gender and Internet experiences and skills (Correa, 2010; Hargittai & Shafer, 2006), we hypothesize that gender inequalities in cultural production in Wikipedia contribution are likely intertwined with Internet experiences and skills. In other words, we ask whether the gender gap in contributions is more, less, or equally present among low and high-skill Internet users.[1] More formally, we test (a) whether an Internet skills divide separates those who do and do not contribute to Wikipedia; and, if so, (b) how this Internet skills divide relates to the gender gap in participation. We conduct these tests using a unique panel survey data set of 547 diverse young adults (ages 21-22 at the time of the second wave) to examine what factors are related to Wikipedia contributions.

We operationalize participation on Wikipedia as any experience making edits to existing entries (including the addition of an image) or starting a new entry. For the purposes of this paper, we do not discriminate between the value of fixing a mistake versus adding new material, editing an existing entry or starting a new one. This is undoubtedly a low threshold for participation, but as we will show, it is meaningful given the low proportion of respondents who engage with Wikipedia in any of those ways.

We use the terms "editor" and "contributor" somewhat interchangeably as the former reflects the common term used within the Wikipedia community to describe anyone who has made

---

[1] For reasons discussed in detail below, we do not test for the presence of a conditional causal process, such as mediation or moderation.

any contribution of any kind to the encyclopedia. Unlike previous studies of the Wikipedia-editing gender gap, our data set includes both individuals who have edited Wikipedia as well as many (indeed, the majority) who have never done so. This diversity makes it possible for us to construct a meaningful test of whether or not Internet experiences and skills may predict whether or not male and female respondents are likely to have ever contributed to Wikipedia.

Our central contribution in this paper consists of an empirical analysis of whether and how Internet experiences and skills may or may not interact with the gender gap in online participation regarding Wikipedia editing. As previous literature on Wikipedia contributions has largely overlooked formal analyses of Internet experiences and skills, the tests we conduct here constitute a novel contribution. Similarly, none of the past work examining Internet skills has focused on contributions to Wikipedia, one of the most widely used online resources and instances of participatory cultural production (ComScore, 2013). In addition, previous work examining gender as well as Internet experiences and skills has not specified the relationship between these variables with much precision. Applying nuanced techniques,we conclude that gender and Internet skills likely have a relatively mild interaction with each other, reinforcing the gender gap at the high end of the Internet skills spectrum.

The rest of the paper is structured as follows: in the next section, we frame our approach and research questions in relation to previous findings on the factors that shape patterns of online participation, and in particular previous work considering the gender gap in Wikipedia contributions. We also briefly review prior work on the role of Internet experiences and skills in explaining variations in a range of online behaviors, including social media participation and other sorts of online cultural production. In the following section, we describe the data set and methods we use to address our research questions, emphasizing the unique characteristics of our sampling frame and longitudinal data collection that make for a particularly compelling test of our research questions.

Then we present the results of our analyses, in which we model the relationship between Wikipedia contributions, gender, Internet experiences and skills, as well as a number of other variables. We end with a discussion of our findings, their limitations, as well as their significance and implications for future work.

What determines who participates in online communities and Wikipedia?

This study contributes to the extensive literature on inequalities in cultural participation specifically focusing on contributions to Wikipedia. Previous studies have emphasized several possible determinants of whether or not individuals contribute to online communities more generally, and to this online encyclopedia more specifically, including background characteristics such as gender as well as Internet experiences and skills. Below, we briefly review related literature.

Background characteristics including gender

Previous work has focused on how certain background characteristics such as gender may shape who participates in online communities. The Wikipedia gender gap represents arguably the most prominent example of this line of inquiry (e.g., Cohen, 2011; Glott et al., 2010). Research into the Wikipedia gender gap originated with the results of a 2008 survey conducted by the Wikimedia Foundation (WMF) and the United Nations University Maastricht (UNU-MERIT) that found that less than 13% of all editors at the time were women, 18% among U.S. adult editors (Glott et al., 2010). Recent analyses taking into account sampling bias introduced by the fact that said survey allowed opt-in participation suggest that these estimates likely overstated the extent of the gender gap whereby 16% of editors globally and 23% in the U.S. are female (Hill & Shaw, 2013). Even the updated figures suggest a considerable gender gap, nonetheless. The presence of such extreme inequality has motivated a growing body of work investigating the relationship between participation in Wikipedia and gender. This work has shown that gender differences extend to editors' earliest patterns of contribution (Antin et al., 2011) and also manifest in how contributors interact with

other editors (Collier & Bear, 2012; Halfaker, Geiger, Morgan, & Riedl, 2012). Related research considers the effects such diversity and differential experiences may have on the quality and quantity of contributions (e.g., Chen, Ren, & Riedl, 2010; Lam et al., 2011) and describes the effects of design interventions intended to reverse the trend (Morgan et al., 2013).

While these previous studies offer insights into gender differences among Wikipedia contributors, the majority of them have sampled exclusively from within the population of Wikipedia editors or used recruitment materials posted on Wikipedia itself. As a result, the respondent pools are biased to include those who already participate or those willing to participate in research announced *through* Wikipedia itself. The results therefore cannot speak to factors that may prevent the majority of females from ever contributing to the encyclopedia in the first place.

Differentiated Internet experiences and skills

Digital inequality research emphasizes that beyond mere physical access to the Internet, people's online experiences also vary based on their autonomy of use, how long they have been Internet users, the amount of time they spend online, their self-efficacy, and their Internet skills (DiMaggio, Hargittai, Celeste, & Shafer, 2004; van Dijk, 2005). People with more autonomy of use, i.e., freedom to use the Internet when and where they wish, engage in more capital-enhancing activities (e.g., Hassani, 2006) as do people who spend more time online (Hargittai & Hinnant, 2008). Self-efficacy (Bandura, 1977) or the belief in one's ability to achieve certain outcomes has also been linked to types of Internet uses (e.g., Eastin & LaRose, 2000; Zhao, Lu, Huang, & Wang, 2010). Research has also found that Web-use skills, i.e., the ability to use the Internet effectively and efficiently, is related to types of online behavior from the likelihood of joining certain communities like Twitter to online content creation and engagement (e.g., Correa, 2010; Hargittai & Hinnant, 2008; Hargittai & Litt, 2011; Wasserman & Richmond-Abbott, 2005). Accordingly, it is important to consider such nuanced measures of Internet experiences when looking at predictors of online

engagement, in this case Wikipedia contributions. While studies of online behavior often include some measure of frequency of use, the other concepts are less likely to show up in analyses, likely in part due to lack of appropriate measures in data sets. Our study overcomes this challenge by drawing on a data set with measures of all of these factors.

Research questions

Our study seeks to understand the relationship of Internet experiences and skills to the act of participating in online cultural production, specifically, contributing to Wikipedia. We ask whether and to what extent Internet experiences and skills – in addition to gender, socioeconomic status, and other background attributes – may help explain who participates in editing Wikipedia. Drawing on the growing literature examining the Wikipedia gender gap as well as past work demonstrating a relationship between online behavior, gender, and skill, we also focus closely on how gender and skill together may account for variation in Wikipedia contribution.

Data and methods

We analyze panel survey data from a diverse group of young adults to look at who contributes to Wikipedia. Selection into the study was completely independent of Wikipedia use thereby avoiding concerns of sampling on the dependent variable. The data set includes a rich set of covariates about both contributors and non-contributors making it possible to examine what factors predict engagement in Wikipedia editing.

We know of no nationally representative panel data set that includes detailed information about people's Internet uses and skills making it impossible to explore these questions on such a sample. Our respondents are diverse on several factors, but homogeneous regarding their age and level of education. These two variables have been shown as important correlates of Internet usage (Dutton and Blank, 2011). By focusing on a generation that was "born digital" (Palfrey and Gasser, 2008) we are excluding people with little-to-no online experiences and thus ensuring some level of

online know-how for the majority of respondents. Accordingly, findings about differences in online participation are likely to be conservative. That is, were we to find effects of Internet experiences and skills on Wikipedia contributions, those are likely to be conservative compared to a more diverse – both demographically and in the related realm of online experiences – sample. We address the implications of the sample's particular make-up in the discussion of our findings.

Data collection

The results reported here draw on two waves of survey data collected about the same people over four years: the first administered in Spring 2009; the second in Summer 2012 on a diverse group of 547 young adults. In Spring 2009, we surveyed a representative sample of first-year students at an urban public university, a school with which neither author has ever had any affiliation beyond this study. Its selection for the project was due to (1) it being one of the most racially and ethnically diverse research universities in the United States (U.S. News & World Report, 2009) relevant for the overall research project's focus on questions of social inequality; and (2) its curriculum requiring that all students take a particular class: the First-Year Writing Program. Working with this course allowed reaching an unbiased sample of the university's student body. It also led to a high response rate since surveys were administered on paper during class sessions.

We used paper-pencil methods as an online questionnaire would have biased against people who are less likely to spend time on the Web, have less private access to the Internet, and who feel less comfortable filling out online forms. For similar reasons, and to control for mode of data collection, the 2012 follow-up study also relied on a paper-pencil instrument.

All 2009 respondents who had consented in 2009 to being recontacted (the majority, 98 percent, had done so) received the 2012 survey in postal mail with a self-addressed stamped envelope. In the first wave of the study, respondents did not receive any incentives. In the 2012

wave, they were offered a $25 gift certificate to Amazon and were entered into a drawing for one of two iPads.

The initial data set in 2009 included valid responses from 1,115 first-year students. The questionnaire included an item to verify respondents' attentiveness to the survey. A small portion, 4.5 percent, answered incorrectly to this verification question, suggesting that they were checking off responses randomly. These students were excluded from the study so as to minimize error. The 1,115 students all answered the verification question correctly. Of the 92 course sections in the program, 86 participated in the study for a 93.5 percent participation rate on the part of sections. Of all the students enrolled in the class, 80.5 percent filled out the questionnaire.

In 2012, we received 547 valid responses. We had permission to follow up with and contact information for 1,066 people for a four percent refusal from the original sample of 1,115. For 17 people, constituting one and a half percent of the original group, both email and postal mail addresses bounced meaning no possibility for recontact.[2] The 2012 questionnaire included two items to verify respondents' attentiveness to question wording. We received 567 surveys, twenty of which – accounting for three and a half percent of the received surveys – had either or both of these questions marked incorrectly or left blank. We excluded these responses from the analyses resulting in 547 valid surveys with data from both years. The 2012 sample is representative of the 2009 sample on most measures regarding background characteristics, Internet experiences and skills. The one statistically significant difference is that a disproportionately smaller number of African Americans participated in the second wave (11% in 2009 vs 8% in 2012).

---

[2] After two attempts at reaching respondents in postal mail, we sent an email to those who had not yet responded to request a verified postal address for mailing the survey in hard copy. No one filled out the survey online.

By 2012, about half of the initial sample was no longer enrolled in college either having graduated (48 percent had done so) or having left school altogether (3.5 percent had done so). For this reason, we refer to respondents as young adults rather than college students.

Measures: independent variables

The survey asked respondents about their demographic and socioeconomic background. We use year of birth to calculate age. We use parental education as a proxy for socioeconomic status, an established practice in the literature (e.g., Hargittai & Litt, 2011). The survey asked about the level of education of both parents using the following categories: (a) less than high school degree; (b) high school degree; (c) some college; (d) college degree (for example: B.A., B.S., B.S.E); (e) advanced graduate (for example: master's, professional, Ph.D., M.D., Ed.D.). Relying on information from these two questions, we created a parental education variable that is assigned the value of the highest education by either parent, e.g., if a student has a mother with a high school degree and a father with a college degree, the parental education variable for that student is coded as "college degree". We then recoded the parental education variable into three mutually exclusive categories: (1) high school degree or less; (2) some college; (3) college degree or more.

To measure race and ethnicity, the survey first asked participants if they were Hispanic or of Latino origin. Then they were asked their race using the following categories: (a) White/Anglo/Caucasian/Middle Eastern; (b) Black/African American; (c) Asian; (d) American Indian or Alaskan Native; (e) Other. Most responses in the "Other" category indicated Hispanic origin and were recoded accordingly. The final categories are: Hispanic, non-Hispanic African American, non-Hispanic Asian American, non-Hispanic Native American, and non-Hispanic White.

To assess how much discretionary time respondents may have, we asked them about their current employment activity. Answer options included "not working" and then several options

specifying amount of weekly work hours. We recoded the responses into a binary variable indicating any time spent working.

In order to gauge the importance of Internet-related experiences for Wikipedia contributions, we collected data on how long people have been Internet users, autonomy of use, and frequency of use. We calculated years of use using information from survey questions that asked about the stage in a student's academic career when he or she first became an Internet user (i.e., during elementary school, middle school, or a particular year in high school). We capped user years at 10 (a response given by just under 20 percent of the sample). To calculate autonomy of use, we relied on questions asking about different locations where participants have access to the Internet (as opposed to actually using it regularly at various locations) by having the choice to check off all applicable locations from a list of eleven options. We created a summary variable from these; the final measure ranges from zero to eleven locations. We logged this measure in the analyses due to the idea that there are diminishing returns to additional locations of access. We derived time spent on the Web weekly (excluding email, chat and voice services) from answers to two questions asking about hours spent on the Web on an average day; one inquiring about weekdays, the other about an average Saturday or Sunday. This measure ranges from 0-42 hours and is also logged in the analyses for reasons similar to logging number of access locations.

To measure Web-use skill, the data set includes a previously-developed and validated instrument included in the Internet Society Module of the General Social Survey 2000 (National Opinion Research Center, 2000; Wasserman & Richmond-Abbott, 2005) although more elaborate and used by others in the literature (e.g., Correa, 2010; Park, 2013). Respondents were asked to rate their level of understanding of 27 Internet-related terms on a five-point scale (Hargittai & Hsieh, 2012). Although a self-report of skill can be prone to gender bias – women tend to rate their skill level lower than men even when controlling for their actual online abilities (Hargittai & Shafer, 2006)

– research has shown that self-perceived skill can itself have implications for what people do online (Hargittai & Walejko 2008) so it is nonetheless a relevant measure. We averaged the 27 items to create the skill index (Cronbach's alpha=.94) with a possible range of 0-4 where the lowest possible score was set to zero.

To get a sense of people's confidence with editing Wikipedia, i.e., self-efficacy (Bandura, 1977), the questionnaire included an item asking respondents to rate how confident they felt in the following statement on a 1-5 point scale (1 referring to "strongly disagree" and 5 meaning "strongly agree"): "I feel confident changing information on a Wikipedia entry."

To account for whether students may have contributed to Wikipedia due to a class assignment, the survey asked the following: "As part of school work, have you ever been assigned to do any of the following?" with "edit an entry on Wikipedia" and "create a new entry on Wikipedia" as answer options. We created a binary variable to signal experience with receiving such an assignment (i.e., a "yes" if the person checked either or both of the above two options).

Measures: dependent variable

To measure contributions to Wikipedia, the instrument asked respondents the following: "How often, if ever, have you done each of the following on Wikipedia?" The activities of interest here are: (1) edited an entry by fixing a mistake; (2) edited an entry by adding new material; (3) started a new entry; and (4) added an image/graphic to an entry. Answer options included "never", "have done it once", "have done it 2-3 times", and "have done it 4 or more times". For the variable measuring Wikipedia contributions, we recoded responses to the above question resulting in a binary measure indicating those who had engaged in any of the four activities ever. The survey also included "read an entry" among the Wikipedia activities, which we use to establish popularity of the site among respondents.

Analyses

First, we present basic descriptive statistics about the popularity of the individual Wikipedia editing activities listed above including details about what proportion of respondents engage more actively on the site. Next, we discuss bivariate descriptive comparisons of how likely people are to contribute to Wikipedia depending on their background characteristics, Internet experiences and skills. Then we run logistic regression analyses to examine what explains differences in likelihood of contributing when controlling for other factors, presenting multiple specifications with and without interaction terms. We also create a graphical representation of the relationship between gender and Internet skills in predicting Wikipedia contribution to highlight the main findings of the regression analyses. For most of the independent variables (background characteristics, Internet experiences and skills, as well as confidence with Wikipedia editing), we use lagged data (data from 2009) to predict experiences with contributing to Wikipedia by 2012. The two independent variables we do not lag are current employment activity and whether the person had ever been assigned to edit Wikipedia for a class.

As previous research on gender and Internet skills suggests either an additive (linear) or multiplicative (interactive) relationship with our outcome variable (editing Wikipedia), we incorporate an interaction term and corresponding hypothesis tests into our regression model (Braumoeller, 2004; Friedrich, 1982). We did not have any basis upon which to anticipate whether the relationship would involve interaction or not (Berr, DeMeritt, & Esarey, 2010; Berry, Rubin, & Esarey, In Press), so we report the marginal effects on the dependent variable using simulation-based estimates of the 95% confidence interval (King, Tomz, & Wittenberg, 2000).[3]

---

[3] This approach follows the procedures recommended by Berry and colleagues (In Press) for circumstances when theoretical support for interaction terms is ambiguous. They demonstrate that normal statistical tests for multiplicative interactions do not possess sufficient statistical power to detect the alternative hypotheses (or reject the null) with conventional levels of precision.

The resulting analysis is similar to statistical tests for "moderation," although we avoid the use of this concept and related terminology of mediation/moderation. The framework of mediation and moderation presumes (1) the presence of a causal process; and (2) that the accompanying statistical techniques can credibly estimate the effects of a causal relationship (Baron & Kenny, 1986; Hayes, 2011). However, following the growing literature on counterfactuals and the potential outcomes framework for causal inference (e.g., Holland 1986; Morgan & Winship, 2007), we contend that such counterfactual outcomes are simply implausible for most analyses involving gender. In the language used by Holland (1986), gender is a stable attribute and therefore cannot provide a suitable basis for credible causal inference using any technique of statistical estimation. Furthermore, the idea that we could test for a causal relationship between gender, Internet skills, and Wikipedia editing is all the more implausible given the nature of our data. Even our detailed, longitudinal survey measures do not provide a credible basis for testing causal claims in the absence of some exogenous intervention or carefully identified instrument (Morgan & Winship, 2007). As a result of these concerns, we adopt a more conservative, careful approach to our analysis and the presentation of our findings. We report the results for multiple model specifications, consider their implications, and use the most compelling specification to build a more intuitive, graphical explanation of the model that also incorporates simulation-based measures of the uncertainty surrounding our estimates (King et al., 2000).

The sample

Everybody in the sample was either 21 or 22 years old so we exclude this variable from the analyses. Women were somewhat more represented than men (see Table 1 for details). Less than half of the sample was White, one-quarter were Asian or Asian American, more than one-fifth were Hispanic, and a little less than a tenth were African American. Just under one-fifth came from families in which neither parent had more than a high school education, just over a fifth had at least

one parent with some college experience, and half had at least one parent with a college degree.

Overall, these figures indicate that the sample is quite diverse from a socioeconomic perspective.

--- TABLE 1 ABOUT HERE ---

In terms of online experiences, on average, these young adults had been using the Internet

for almost six years at the time of the first wave of data collection (2009), had numerous locations to

access the Web, and spent approximately 20 hours per week online (not counting email, chat, or

voice services). The average respondent's confidence in editing Wikipedia is relatively low, falling

between the "disagree" and "neutral" sentiments regarding the confidence statement, but tending

toward the former. About one in eight students had been given an assignment in class at some point

either to edit or create a new entry on Wikipedia. While the sample consists of a highly wired group

of young adults, we observe considerable variation in Internet skills, ranging from 0.33 to 4 on a 0-4

scale. See Table 2 for details.

--- TABLE 2 ABOUT HERE ---

Popularity of contributing to Wikipedia

The vast majority (99%) of respondents reported having read an entry on Wikipedia,

signaling that lack of contributions cannot be attributed to not knowing about or not visiting the

site. Looking at aggregate statistics of how many respondents made edits to existing entries

(including the addition of an image) or started a new entry, we find that just over a quarter (28%) of

the sample reports such experiences (many having done this just once ever). This figure drops to just

a fifth of respondents if we exclude those who had been assigned to make edits to Wikipedia as part

of a class assignment.

As Table 3 shows, the most popular engagement is editing an entry by fixing a mistake,

which just under a quarter of the sample reported having done at least once. Less than a fifth

claimed experiences with adding new material while less than a tenth had either started a new entry

or had added an image or a graphic to an entry, respectively. Overall, these figures suggest that contributing to Wikipedia is not a widespread activity. On the whole, sixteen percent of the sample reported having ever engaged in more than one of the four activities and only twelve percent had engaged in any of the activities more than once (less than five percent had engaged in any of the activities 4 or more times).

--- TABLE 3 ABOUT HERE ---

Wikipedia contributions by user attributes

Table 4 presents what percentage of sample subgroups reported any Wikipedia contributions. Consistent with the literature, women are much less likely to have contributed to Wikipedia than men. Only one in five females (21%) ever made any Wikipedia edits compared to over a third of men (38%). The discrepancies are even larger when we only look at those who did not have a related class assignment: 13% of such women compared to almost 32% of such men made some edits to the site. The observed differences in any level of engagement also extend to intensity of engagement. Men are much more likely to have engaged more than once in Wikipedia edits than women, a result especially salient for fixing a mistake and adding new material to an entry. Generally speaking, very few participants had started new entries or added images to entries more than once regardless of their gender.

Additional relationships reported in Table 4 suggest that Whites are more likely to edit than others as are those from more educated families. These are the only background characteristics that exhibit statistically significant relationships with Wikipedia contributions. Various Internet-use variables do matter considerably, however. Because these measures are continuous, we created quartiles and present the percent of contributors for the lowest (LQ) and highest (HQ) quartile groups of each measure. Frequent Web users are more likely to contribute as are those who are more confident in their Wikipedia editing abilities, both concerning lagged measures. The biggest

contrast between contributors and non-contributors concerns general Internet skills as a

considerably larger segment of those in the highest skill quartile have made at least one edit on

Wikipedia compared to those in the lowest quartile. Next we look at whether these findings hold

when considering people's background characteristics and Internet experiences simultaneously.

--- TABLE 4 ABOUT HERE ---

Predicting Wikipedia contributions

The results presented in Table 5 of logistic regression analyses show evidence that both

gender and skill explain variation in Wikipedia contributions in several model specifications.

Remarkably, these two variables – along with the measure of whether or not respondents had been

assigned to edit Wikipedia for a class – are the only significant predictors of Wikipedia contributions

when holding other factors constant.[4] None of the other variables in the models – including

measures of race, ethnicity, socioeconomic status, availability, Internet experience, and confidence in

editing Wikipedia – are significant.

Consistent with previous research on who edits Wikipedia, female respondents were less

likely than males to have ever contributed to the encyclopedia across most specifications. Adding to

this literature is our novel introduction of Internet skills in the model, a variable that is a significant

predictor of editing as well, even while controlling for gender. Individuals with high Internet skills

were considerably more likely to contribute than those with low Internet skills. We note that the

addition of Internet skills attenuates the relationship between gender and Wikipedia contribution

(Models 3 and 5).

Models 4 and 6 introduce the multiplicative term capturing the interaction of gender and

Internet skills, and both show that the interaction term itself is not a significant predictor of editing

---

[4] All results reported here are robust to the exclusion of the subset of respondents who had been assigned to edit Wikipedia for a class. In additional robustness checks, we included a control for whether or not respondents had ever contributed to Wikipedia by 2009, and our findings were unchanged substantively.

Wikipedia. In Models 3 and 4, the Internet skills measure runs from 0-4, whereas in Models 5 and 6 we present results with Internet skills centered around its mean value. As expected, these models are identical except for the coefficients for the Intercept and Female. The differences between these specifications illustrate the nature of the relationship between gender and Internet skills and their joint association with Wikipedia editing. The addition of the interaction term in Model 4 attenuates the association between gender and Wikipedia contribution completely. However, centering Internet skills around the mean recovers the significant relationship between gender and the dependent variable.[5] The interaction term for gender and Internet skills is not significant, but this provides inconclusive evidence for the presence or absence of an interactive relationship (Berry et al. In Press).

We investigate this relationship between gender and Internet skills further in a plot of marginal effects in Figure 1. All covariates other than gender and Internet skills are held at their sample mean values (modal values for categorical predictors) and the coefficients from Model 6 (Table 5) are used to estimate predicted probabilities that a male or female respondent at each observed point on the Internet skills index would have edited Wikipedia.[6] In addition, "rug plots" (ticks along the top and bottom of the horizontal axes of the figure) show the distribution of male and female respondents respectively along the Internet skills index. The grey "ribbons" accompanying the two curves represent the 95% confidence interval around these predictions, which we generated by simulating 5000 randomly drawn parameter values with identical variance-covariance matrices to the observed model from a multivariate normal distribution (King et al, 2000). The divergence of the curves further illustrates the nature of the interaction: at low values for Internet skills, the predicted probabilities of Wikipedia contribution for males and females overlap

---

[5] This is due to the fact that, in the interaction models, the coefficient for gender only captures the relationship with the dependent variable when the other component of the interaction term (Internet skills) is equal to zero.

[6] The figure is substantively the same if we use the coefficients from Model 4.

almost perfectly. As the observed level of Internet skills increases, the curves drift apart and the

accompanying confidence intervals nearly separate. We discuss these results in greater detail below.

--- FIGURE 1 ABOUT HERE ---

Finally, Figure 2 visualizes the interactive relationship between gender, Internet skills, and

Wikipedia editing estimated by Model 6 in greater detail.[7] Here the test statistics for the model

coefficient for gender are visualized at discrete points along the distribution of Internet skills and the

critical value for the test statistic ($|t| > 1.97$) is indicated by the red horizontal segment. The plot

shows that for values of Internet skills roughly below the sample mean (2.21), the model cannot

reject the null hypothesis of no effect for gender. In other words, below the sample mean for

Internet skills we find no support for the hypothesis of a relationship between gender and Internet

skills, but above it we do. These results are suggestive of an interaction between gender and Internet

skills, but remain inconclusive for reasons addressed by Berry and colleagues (In Press).

--- FIGURE 2 ABOUT HERE ---

Discussion and conclusion

Our findings suggest that Internet skills are an extraordinarily robust predictor of

contributing to Wikipedia and also help explain a key dimension of the gender gap among Wikipedia

editors. Using data that includes detailed and reliable measures of demographic attributes as well as

socioeconomic status, we find that only gender and Internet skills help explain variations in

Wikipedia contribution when excluding related school assignments. Most importantly, we find that

the variation in Internet skills (measured in 2009) and gender create a consistent pattern in terms of

who contributes to Wikipedia years later (by 2012). The models also suggest the presence of a

substantively meaningful interaction between gender and Internet skills, despite failing to reject the

null hypothesis of no association for the multiplicative term itself. As the marginal-effects plot

---

[7] Again, the figure is unchanged if we use Model 4.

presented in Figure 1 illustrates, we find that higher levels of Internet skills predict much greater probability of contribution for men than for women. Given that women tend, on average, to have lower Internet skills than men, this indicates a key reason why the gender gap may be so persistent.

The results also suggest that low-skilled men and low-skilled women are equally highly *unlikely* to contribute to Wikipedia. In other words, there is no gender gap in Wikipedia contribution among individuals with low Internet skills. The fact that we see these results in a panel data set where we surveyed the same individuals years apart suggests that differentiated Internet know-how has long-term consequences. Women in the study have lower skills than comparable men, helping to explain why they are so much less likely ever to have edited Wikipedia than their male counterparts.

Why women, on average, report lower-level understanding of Internet-related terms remains a puzzle. Although studies with detailed data about actual skills based on performance tests suggest no gender differences in the observed skills (Hargittai & Shafer, 2006; van Deursen & van Dijk, 2010), research that looks at self-rated know-how consistently finds gender variation (Correa, 2010; Wasserman & Richmond-Abbott, 2005) with real consequences for online behavior (e.g., differentiated rates of content sharing online [Hargittai & Walejko, 2008]). That is, even if it is just a perceived variation in know-how, it has real consequences for online behavior as demonstrated in the present study as well. One possible source of this discrepancy could be confidence in using the Internet, but as noted above, we control for confidence with the specific activity of editing Wikipedia and that does not seem to make a difference for actual editing experiences. Another possible source of this variation concerns the level of encouragement, or lack thereof, that boys versus girls receive for tinkering with technology while growing up. If parents place more restrictions on girls' technology uses out of safety concerns, that may scare girls away from open exploration with potential long-term consequences for understanding the ins-and-outs of the

medium. Future research should tackle such potential negative effects of restrictions on technology use during childhood.

This study offers novel contributions and also opens up new arenas of inquiry for research on social media, peer production, online behavior more broadly and the general question of cultural production in the digital age. Given the type of data we have, we are unable to differentiate among various types of "lurkers" for example, people who visit Wikipedia, but have not contributed in the ways we measure participation (i.e., specific content additions to entries). Scholars have debated how being a member of the audience is itself a type of participation (Crawford, 2009; Jenkins, Ford, & Green, 2013; Van Dijck, 2009), an area of inquiry that offers interesting questions for future research that will require different data to address.

Having measured Wikipedia editing experiences through class assignments, we are able to show that educational interventions can matter to increasing participation. Our project did not focus on such interventions, but they do exist (e.g., FemTechNet) and are ripe area for research to see whether they can help alleviate gender and skill differences in Wikipedia contributions.

We extend earlier work on the Wikipedia gender gap by demonstrating that variation in Internet skills helps explain a critical aspect of the variation in the contribution rates among men and women. Our novel finding that the gender gap in Wikipedia contributions does not exist for individuals with low Internet skills complicates prior explanations of these phenomena. While studies have demonstrated differential participation practices among Wikipedia editors (e.g., Antin et al., 2011; Kriplean, Beschastnick, & McDonald, 2008; McDonald, Javanmardi, & Zachry, 2011; Welser et al., 2011), we are not aware of work offering insights into specific factors that explain how people select into contributing to Wikipedia in the first place, beyond gender. We are able to make these unique contributions, because we have comparable data on both contributors and non-

contributors thereby avoiding a common limitation of prior work: sampling on the dependent variable, i.e., analyzing data only about contributors.

The fact that Internet skills are such a robust predictor – at least as robust as gender – of who participates in Wikipedia editing underscores the importance of incorporating skill as both an object of future research studies and as a central design consideration for social media and peer production systems. For example, much work has focused on the motivations driving participation in online communities and on empirically testing design interventions that effectively incentivize participants to contribute (e.g., Kraut & Resnick, 2011). These are important dimensions of building successful, dynamic online communities, but interventions must also address differentiated levels of Internet skills.

One limitation of this study, the fact that it examines a nonrandom and statistically non-representative sample of young adults, restricts our ability to generalize on the basis of these findings. Nevertheless, panel data with detailed measures of Internet skills and use remain extremely rare and are non-existent for nationally representative samples. Thus data even on non-representative samples can offer unique insights into the social dynamics and determinants of participation in online communities. In addition, the fact that this data set incorporates many non-contributors to Wikipedia allows us to explore the factors that predict contribution in the first place. Moreover, given that respondents represent a highly-wired panel of young adults (i.e., not novices when it comes to digital media), the findings are likely conservative when it comes to the importance of skill among a more diverse group.

A great deal of work on these topics remains to be done. How do gender and skill differences in contribution manifest themselves among people heterogeneous in age and education? Why is it that women with high levels of Internet skills are so much less likely to contribute to Wikipedia than comparably-skilled men? While research has shown that one of the reasons women

use the Internet differently from men concerns their parental responsibilities due to less available time (Kennedy, Wellman, & Klement, 2003), given that only three percent of the sample reported having a child, this factor cannot explain the gender differences for this group. Self-selection into a project like Wikipedia also plays a critical role, but specific reasons why women might find encyclopedia writing (or the practice of doing so in Wikipedia) disproportionately unappealing requires further research.

Additionally, future research should seek to explain the pathways through which Internet skills relate to participation in Wikipedia and comparable online communities. More work is needed in understanding what particular Web-use skills – not necessarily specific to Wikipedia – are especially helpful for encouraging contributions and what types of interventions can improve them, especially among women. In addition, the relationship between skills, gender, and other predictors of participation should be examined in greater depth through multiple methods. Our findings partially contradict previous studies of the Wikipedia gender gap for participation and begin to unpack the mechanisms by which Internet skills (along with gender and other factors) contribute to differentiated participation in cultural production online.

References

Alderson, A.S., Junisbai A., & Heacock A. (2007) Social status and cultural consumption in the United States. *Poetic,s* 35:191-212.

Perrin, A., & Vaisey, S. (2008) Parallel Public Spheres: Distance and Discourse in Letters to the Editor. *American Journal of Sociology,* 114:781-810. DOI: 10.1086/590647.

Antin, J., Yee R., Cheshire C., & Nov O. (2011) Gender differences in Wikipedia editing, *ACM WikiSym*, Mountain View, California. pp. 11-14.

Bandura, A. (1977) Self-Efficacy:Toward a Unifying Theory of Behavioral Change. Psychological Review 84:191-215.

Baron, R. M., & Kenny, D. A.D. (1986) The Moderator-Mediator Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51: 1173-1182.

Benkler, Y. (2002) *Coase's penguin, or, Linux and the nature of the firm*. Yale Law Journal 112:369-446.

Benkler, Y. (2006) *The Wealth of Networks*. Yale University Press, New Haven, CT.

Berry, W.D., DeMeritt J.H.R., & Esarey J. (2010) Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential? *American Journal of Political Science* ,54:248–266.

Berry, W.D., Rubin J.H., & Esarey J. (In Press) Bias and Overconfidence in Parametric Models of Interactive Processes. *American Journal of Political Science.*

Blank, G. (2013) Who Creates Content? Stratification and Content Creation on the Internet. *Information, Communication & Society,* 16:590-612. DOI: 10.1080/1369118x.2013.777758.

Braumoeller, B.F. (2004) Hypothesis Testing and Multiplicative Interaction Terms. *International Organization*, 58:807–820.

Chen, J., Ren, Y., Riedl, J. (2010) The effects of diversity on group productivity and member withdrawal in online volunteer groups. *Proceedings of the 28th international conference on Human factors in computing systems, ACM*, New York, NY, USA. pp. 821–830

Cohen, N. (2011a) Define Gender Gap? Look up Wikipedia's Contributor List. *The New York Times*, New York.

Collier, B., & Bear, J. (2012) Conflict, Criticism, or Confidence: An Empirical Examination of the Gender Gap in Wikipedia Contributions. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, ACM*. pp. 383-392.

ComScore. (2013) comScore Media Metrix Ranks Top 50 U.S. Web Properties for May 2013. Reston, Virginia.

Correa, T. (2010) The Participation Divide Among "Online Experts": Experience, Skills and Psychological Factors as Predictors of College Students' Web Content Creation. *Journal of Computer-Mediated Communication*, 16:71-92. DOI: 10.1111/j.1083-6101.2010.01532.x.

Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum: Journal of Media & Cultural Studies*, 23(4), 525-535.

DiMaggio, P., & Useem, M. (1978) Social Class and Arts Consumption: The Origins and Consequences of Class Differences in Exposure to the Arts in America. *Theory and Society* ,5:141-161.

DiMaggio, P., & Ostrower, F. (1990) Participation in the Arts by Black and White Americans. *Social Forces*, 68:753-778.

DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004) Digital Inequality: From Unequal Access to Differentiated Use, in: K. Neckerman (Ed.), *Social Inequality.* Russell Sage Foundation, New York. pp. 355-400.

Dutton, W.H., & Blank, G. (2011) Next Generation Users: The Internet in Britain. Oxford Internet

    Survey 2011. Oxford Internet Institute, University of Oxford.

Eastin, M.S., & LaRose, R. (2000) Internet Self-Efficacy and the Psychology of the Digital Divide.

    Journal of Computer Mediated Communication, 6. DOI: 10.1111/j.1083-6101.2000.tb00110.x.

Ecker,t S., & Steiner, L. (2013) (Re)triggering Backlash: Responses to News About Wikipedia's

    Gender Gap. *Journal of Communication Inquiry* ,37:284-303. DOI: 10.1177/0196859913505618.

Filipacchi, A. (2013a) Wikipedia's Sexism, The New York Times.

Friedrich, R.J. (1982) In Defense of Multiplicative Terms in Multiple Regression Equations. *American

    Journal of Political Science* ,26:797-833.

Gleick, J. (April 29, 2013) Wikipedia's Women Problem. *The New York Review of Books*, p. SR3

Glott, R., Ghosh R., & Schmidt P. (2010) Wikipedia Survey, UNU-MERIT, Maastricht,

    Netherlands.

Hayes, A. (2013) *Introduction to Mediation, Moderation, and Conditional Process Analysis.* The Guilford

    Press, New York.

Halfaker, A., Kittur, A., & Riedl, J. (2011). Don't bite the newbies: how reverts affect the quantity

    and quality of Wikipedia work. Paper presented at the ACM WikiSym Mountain View,

    California.

Halfaker, A., Geiger R.S., Morgan J.T., & Riedl J. (2012) The Rise and Decline of an Open

    Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline.

    American Behavioral Scientist ,57:664-688 DOI: 10.1177/0002764212469365.

Hargittai, E. (2000). Open Portals or Closed Gates? Channeling Content on the World Wide Web.

    Poetics. 27(4), 233-254.

Hargittai, E., & Shafer, S. (2006) Differences in Actual and Perceived Online Skills: The Role of

    Gender. *Social Science Quarterly.* 87(2), 432-448

Hargittai, E., & Walejko, G. (2008) The Participation Divide: Content Creation and Sharing in the
        Digital Age. *Information, Communication and Society* ,11:239-256.

Hargittai, E., & Hinnant, A. (2008) Digital Inequality: Differences in Young Adults' Use of the
        Internet. *Communication Research,* 35:602-621.

Hargittai, E., & Litt, E. (2011) The tweet smell of celebrity success: Explaining variation in Twitter
        adoption among a diverse group of young adults. *New Media & Society*, 13(5), 824-842. DOI:
        10.1177/1461444811405805.

Hargittai, E., & Hsieh, Y. P. (2012). Succinct survey measures of Web-use skills. Social Science
        Computer Review, 30(1), 95-107.

Hargittai, E., & Hsieh, Y. P. (2013). Digital Inequality. In W. H. Dutton (Ed.), *Oxford Handbook for
        Internet Research,* (pp. 129-150.) Oxford, England: Oxford University Press.

Hassani, S.N. (2006) Locating digital divides at home, work, and everywhere else. *Poetics,* 34:250-272.

Hill, B.M., & Shaw, A. (2013). The Wikipedia Gender Gap Revisited: Characterizing Survey
        Response Bias with Propensity Score Estimation. PLOS ONE, 8(6), e65782. doi:
        10.1371/journal.pone.0065782

Holland, P. (1986) Statistics and Causal Inference. *Journal of the American Statistical Association,* 81(396):
        945-960.

Howard, P.N., Rainie, L., & Jones, S. (2001) Days and Nights on the Internet: The Impact of a
        Diffusing Technology. *American Behavioral Scientist,* 45:383-404.

Jenkins, H. (2006) *Convergence culture: where old and new media collide*. New York University Press New
        York.

Jenkins, H., Ford, S., & Greene, J. (2013) *Spreadable Media: Creating Value and Meaning in a Networked
        Culture*. NYU Press, New York.

Kennedy, T., Wellman, B., & Klement, D. (2003) Gendering the Digital Divide. *IT & Society* 1.

King, G., Tomz, M., & Wittenberg, J. (2000) Making the Most of Statistical Analyses: Improving

>   Interpretation and Presentation. *American Journal of Political Science,* 44: 341-355.

Kraut, R.E., & Resnick, P. (2011) *Building Successful Online Communities: Evidence-Based Social Design.*

>   MIT Press, Cambridge, MA.

Kriplean, T., Beschastnikh, I., & McDonald, D.W. (2008) Articulations of wikiwork: uncovering

>   valued work in wikipedia through barnstars. *ACM CSCW '08*, pp. 47-56.

Lam S.K., Uduwage A., Dong Z., Sen S., Musicant D.R., Terveen L., & Riedl J. (2011)

>   WP:Clubhouse?: An Exploration of Wikipedia's Gender Imbalance. *ACM WikiSym '11*, pp.

>   1-10.

McDonald, D.W., Javanmardi, S., & Zachry, M. (2011) Finding patterns in behavioral observations

>   by automatically labeling forms of wikiwork in Barnstars, ACM WikiSym '11. pp. 15-24.

Morgan, J.T., Bouterse, S., Walls, H., & Stierch, S. (2013) Tea and sympathy: crafting positive new

>   user experiences on Wikipedia. *ACM Computer Supported Cooperative Work*, San Antonio,

>   Texas. pp. 839-848.

Morgan, S. L., & Winship, C. (2007) *Counterfactuals and Causal Inference: Methods and Principles for Social*

>   *Research.* Cambridge University Press, New York.

National Opinion Research Center. (2000) General Social Survey Information Society Module.

>   Chicago, IL.

Palfrey, J., & Gasser, U. (2008) *Born Digital: Understanding the First Generation of Digital Natives.* Basic

>   Books, New York.

Park, Y.J. (2013) Digital Literacy and Privacy Behavior Online. *Communication Research,* 40:215-236.

>   DOI: 10.1177/0093650211418338.

Schradie, J. (2011) The digital production gap: The digital divide and Web 2.0 collide. *Poetics,* 39:145-

>   168.

U.S. News & World Report. (2009) America's Best College 2010. Washington, DC.

van Deursen, A., & van Dijk, J. (2010). Measuring Internet Skills. International Journal of Human-Computer Interaction, 26(10), 891-916.

van Dijck, J. (2009) Users like you? Theorizing agency in user-generated content. *Media Culture and Society*, 31, 41-58.

van Dijk, J.A.G.M. (2005) The Deepening Divide Sage Publications, Inc., Thousand Oaks, CA.

Wasserman, I.M., Richmond-Abbott, M. (2005) Gender and the Internet: Causes of Variation in Access, Level, and Scope of Use. *Social Science Quarterly,* 86:252-270.

Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011) Finding social roles in Wikipedia. *iConference Fort Worth*, Texas. pp. 122-129.

Zhao, L., Lu, Y., Huang, W., & Wang, Q. (2010) Internet inequality: The relationship between high school students' Internet use in different locations and their Internet self-efficacy. *Computers & Education,* 55:1405-1423. DOI: DOI: 10.1016/j.compedu.2010.05.010.

Zickuhr, K., & Smith, A. (2012) Digital Differences, Pew Internet & American Life Project, Washington, D.C.

Zillien, N., & Hargittai, E. (2009). Digital Distinction: Status-Specific Internet Uses. *Social Science Quarterly, 90*(2), 274-291.

User attributes	Percent
Gender	
Men	39
Women	61
Race/ethnicity	
African American, non-Hispanic	8
Asian American, non-Hispanic	24
Hispanic	23
Native American, non-Hispanic	1
White, non-Hispanic	44
Parental education	
High school or less	23
Some college	26
College or more	50
Employment activity	
Currently working	76
Currently not working	24

Table 1. Sample descriptives

Internet experiences	Mean	Standard deviation
Number of use years	5.71	2.25
Number of Internet access locations	6.48	2.18
Hours spent on the Web weekly	17.16	9.76
Confidence in editing Wikipedia (1-5)	2.17	1.14
School assignment to edit Wikipedia (2012 data)	.12	.33
Internet skills (0-4)	2.2	.82

Table 2. Internet experiences (2009 data)

Wikipedia editing experiences	Never	Has done it once	Has done it 2-3 times	Has done it 4 or more times
Read an entry	1.10	1.29	5.71	91.90
Edited an entry by fixing a mistake	75.87	12.89	7.18	4.05
Edited an entry by adding new material	82.69	10.50	3.68	3.13
Added an image/graphic to an entry	90.81	5.33	1.84	2.04
Started a new entry	91.18	5.51	1.10	2.21

Table 3. Level of engagement with Wikipedia editing activities

User attributes	Full Sample	Excluding those with a related school assignment
Gender		
Men	38.21***	31.55***
Women	20.90***	12.97***
Race/ethnicity		
African American, non-Hispanic	20.93	15.38
Asian American, non-Hispanic	23.44	20.17
Hispanic	25.62	14.71
White, non-Hispanic	31.65*	23.90*
Parental education		
High school or less	25.20	14.81
Some college	28.37	18.85
College or more	28.57	23.36*
Employment activity		
Currently working	29.09	21.70
Currently not working	23.08	15.65
Internet experiences		
Number of use years lowest quartile (LQ)	14.29	16.67
Number of use years highest quartile (HQ)	29.35	21.41
Autonomy of use LQ	24.04*	17.39
Autonomy of use HQ	33.33*	25.15
Frequency of use LQ	21.21*	11.76*
Frequency of use HQ	33.11*	26.52*
Confidence in editing Wikipedia Lowest Third	24.49***	15.29***
Confidence in editing Wikipedia Highest Third	37.56***	31.18***
Internet skills LQ	14.73***	7.76***
Internet skills HQ	39.86***	32.50***

Table 4. Wikipedia contributions by user attribute, prior Internet experiences and skills (asterisks signify statistically significant differences between the subgroups of the category; *p<.1 *** p<.001)

Table 5. Logistic regression on contribution to Wikipedia (coefficients reported as log-odds with standard errors in parentheses)

	1	2	3	4	5	6
Background						
Female	-0.90***	-0.91***	-0.71**	0.00	-0.71**	-0.66*
	(0.20)	(0.24)	(0.26)	(0.69)	(0.26)	(0.27)
Hispanic	-0.13	-0.19	-0.20	-0.21	-0.20	-0.21
	(0.28)	(0.34)	(0.34)	(0.35)	(0.34)	(0.35)
African American non-Hispanic	-0.37	-0.33	-0.44	-0.40	-0.44	-0.40
	(0.41)	(0.48)	(0.48)	(0.48)	(0.48)	(0.48)
Asian American non-Hispanic	-0.35	-0.28	-0.39	-0.37	-0.39	-0.37
	(0.26)	(0.29)	(0.30)	(0.30)	(0.30)	(0.30)
Parental Education Some College	-0.07	-0.08	-0.04	-0.04	-0.04	-0.04
	(0.25)	(0.28)	(0.29)	(0.29)	(0.29)	(0.29)
Parental Education High School or less	-0.18	-0.23	-0.17	-0.17	-0.17	-0.17
	(0.28)	(0.33)	(0.34)	(0.34)	(0.34)	(0.34)
Working (2012)	0.31	0.32	0.32	0.31	0.32	0.31
	(0.25)	(0.29)	(0.29)	(0.29)	(0.29)	(0.29)
Internet experiences						
Internet use years (2009)		0.06	0.03	0.03	0.03	0.03
		(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Internet autonomy (2009)		0.14	-0.04	-0.03	-0.04	-0.03
		(0.34)	(0.34)	(0.34)	(0.34)	(0.34)
Internet use frequency (2009)		0.40	0.31	0.30	0.31	0.30
		(0.20)	(0.21)	(0.21)	(0.21)	(0.21)
School assignment to edit Wikipedia (2012)		3.00***	2.97***	2.98***	2.97***	2.98***
		(0.36)	(0.36)	(0.36)	(0.36)	(0.36)
Confidence editing Wikipedia (2009)		0.16	0.06	0.04	0.06	0.04
		(0.10)	(0.11)	(0.11)	(0.11)	(0.11)
Internet skills (2009)			0.52**	0.69**	0.52**	0.69**
			(0.17)	(0.23)	(0.17)	(0.23)
Female*Internet skills				-0.35		-0.35
				(0.31)		(0.31)
(Intercept)	-0.51	-3.03**	-3.16**	-3.50***	-2.18*	-2.20*
	(0.29)	(0.96)	(0.96)	(1.02)	(0.99)	(1.00)
AIC	612.20	510.84	503.00	503.75	503.00	503.75
BIC	646.38	566.14	562.53	567.53	562.53	567.53
Log Likelihood	-298.10	-242.42	-237.50	-236.88	-237.50	-236.88
Deviance	596.20	484.84	475.00	473.75	475.00	473.75
Num. obs.	530	520	519	519	519	519

***$p<0.001$,**$p<0.01$,*$p<0.05$

Model 1 includes background variables only; Model 2 adds Internet experiences variables; Model 3 adds Internet skills; Model 4 adds an interaction of Internet skills & Female; Models 5 & 6 use Internet skills centered at the sample mean.
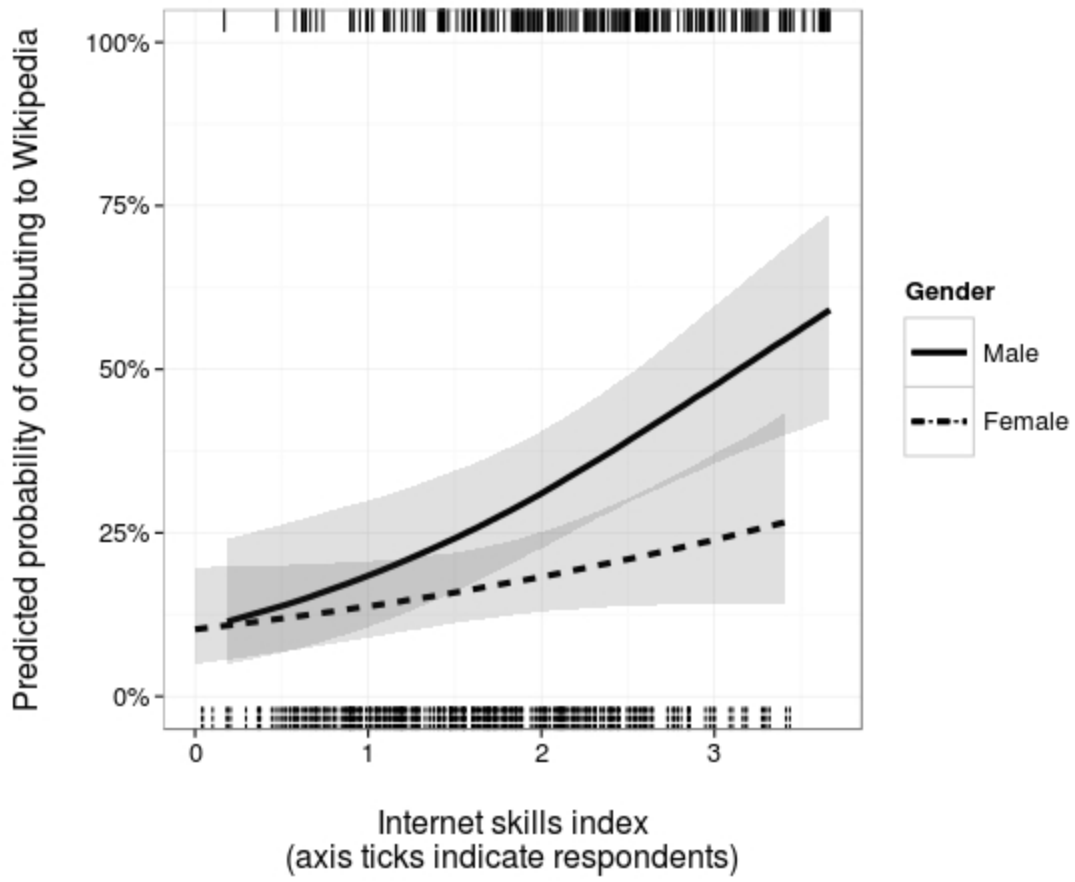
Figure 1. Marginal effects plot: Predicted probability of contribution to Wikipedia by gender
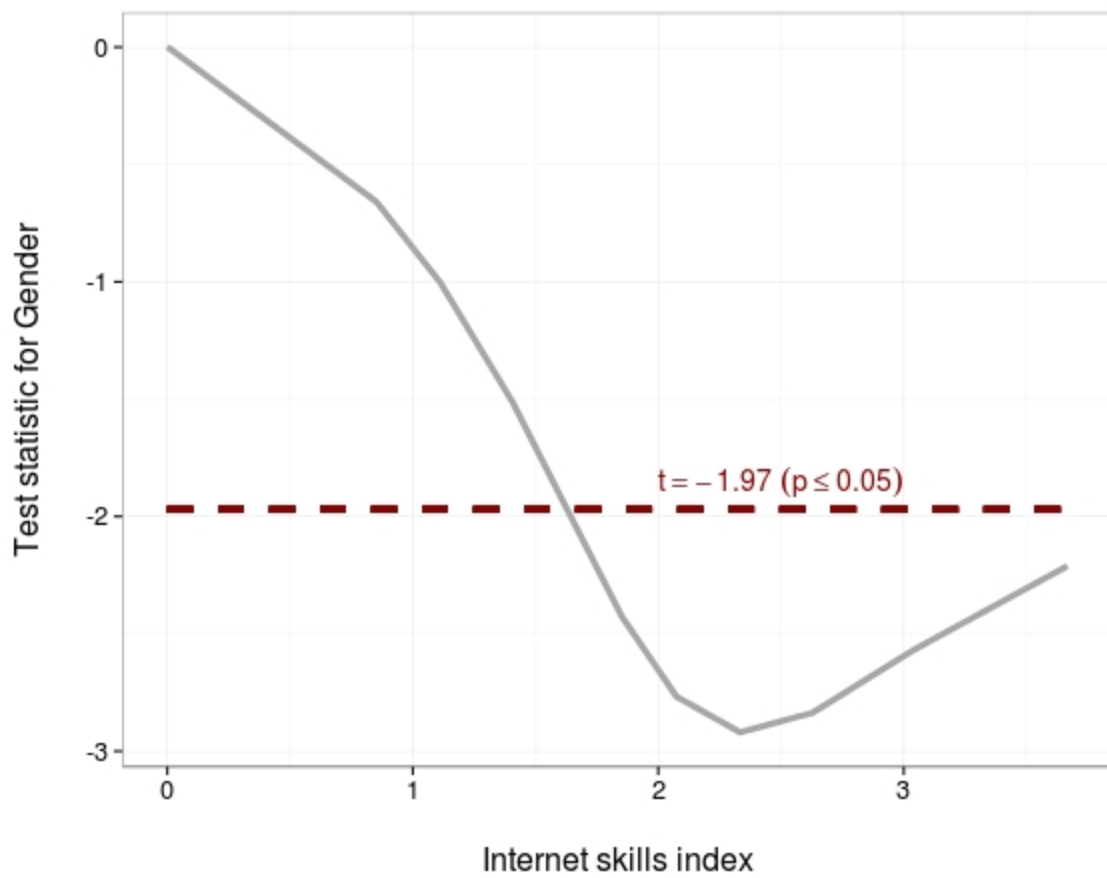
Figure 2. Test statistic for the relationship between gender and Wikipedia editing across the range of observed values for Internet skills