

*These transcriptions may contain errors, especially in spelling of names. These are unfortunate, and we regret that we do not have the resources to fix these errors. Still we believe these transcripts will be valuable to many users.*

## **The Implications of Digital Inequality for Internet Research**

### **Eszter Hargittai**

>> Eszter Hargittai: My goal here today is to focus on and point out how much people's background matters and how they use the internet, and then focus on a few specific aspects of that question, and then talk about how this relates to various research design issues that you may be implementing in your studies. The overall framework that I work with is, I start out with users of digital media. So the traditional digital design research looks at access versus no access, question and...that actually remains a very important question, it's just not what I study today. I study users and differences among them, recognizing, very explicitly, that people come to their uses from a certain socioeconomic background, and then also focus on the context of their uses, both technical and social. So not assuming that people have the same context in which they use digital media but that those factors differ. And then what I've argued for over a decade is that people's skills will differ when it comes to using the internet. I'll talk more about what I mean by that. And then the argument is that skills and all those other factors will influence what people do online both in terms of content consumption and content production, participation, active involvement. And then the really big question of interest, especially to scholars of social stratification, but I would argue to lots of other scholars as well is, to make the argument that the internet matters, we need to know whether all these different types of online activities translate into life outcomes like better well-being, better health outcomes, more productivity, better labor markets, outcomes, etc. Those are lots of outcomes that one could study and where the internet might matter, but we have lots of work to do in figuring out how it might matter. As I said, skill is something I've focused on for over a decade in my work, and more generally speaking, what I'm interested in is, who are the people who benefit from their digital media uses? So just taking one more step back, in terms of sociological theory...and I'm a sociologist by training, although now I'm in our communications studies department. And in terms of professionalization, if anyone's interested in talking about what it's like to switch fields like that, feel free to grab me later. So the question is, whether use of digital media will allow people to improve their life chances, that would be social mobility, or whether where people are in society pretty much just gets reproduced regardless of their digital media uses. So those are the big overarching questions driving this research, arguing that just because you're online, itself, isn't going to mean that you're going to improve your life chances. And then, I could talk about this later in more detail, but what are the processes by which people's life chances might improve? These are some of those that are...whether it's...I don't think, to this crowd, I really need to explain this, but your internet uses could improve your various skills that could help you get a better job, improving your financial capital, etc. I can get into detail in this later as well. There's been lots of research over the years about social inequality. I'm certainly not the only one to do research on this, and it's been done across countries. Here are some references. Patrick and I wrote a review of digital inequality research that's coming out soon, which has lots of details about this issue. And pretty consistently, people have found, for different national contexts, that socioeconomic status relates to various information seeking activities; however, very little of this work looks at skill, which I argue is one of the processes by which these relationships exist, and also, less work looks at more participatory activity. So those are some of the things that I've been doing in my work. So as I've said, skill is something I've focused on a lot, so what do I mean by that? That could mean a lot of things. At a very core basic level, I consider it to mean awareness and understanding of what's possible online and what we can do, even just sort of people's cognitive map of what they can go online to do. I think for those of us who research digital media and think about digital media all the time and spend a lot of time online, it's very hard to recognize that to lots of people it's not as obvious that you turn online to answer various questions, whatnot. So that's a very basic level of skill. And then there's the issue of efficient information seeking, and

I've done a lot of work on that, being able to find information efficiently, effectively. Also, being able to evaluate the credibility of the content that you find, this is another huge area, and, in fact, an increasing area with all the misinformation that increasingly gets out there in all sorts of realms, whether that's politics or health or whatnot. And then also, I think, increasing over time has been more and more focus and concern about privacy and security issues and to what extent do people recognize the risks in that realm online. And then related to all this, as I will talk about, is how much are people participating. And one of the big questions, and we've heard other people already discuss this at this workshop, is figuring out why people contribute to various communities and online activities. It's a huge question. It's a difficult question. My argument is part of it has to do with skill. So even some people might want to, but they might not know how to. And then the question is, do they have the resources to get help with that to be able to contribute and have their voices heard. So what do you do if you're studying something or interested in something like internet skill that no one else, certainly at the point when I started doing this over a decade ago, had data on. Well you end up collecting your own data. So I'm going to go back a decade, and this is work that I did for my dissertation. And as Jenny mentioned, I was involved with Webshop, and I came down to Maryland a few times those few years. And Ben was super supportive over this work, which I really appreciate very much. This was a pretty unorthodox dissertation to be doing in a sociology department, so most were very critical to my advisor, Paul DiMaggio, who very much supported this. It was fantastic to come here. And, Ben, I can still remember just Ben being so excited seeing my work, so that was really, really encouraging. But what I did, just in a few words, I brought in people into a lab and sat with them one on one, asked them to perform various tasks in different realms and basically came up with measures of their actual skill, which is pretty much...it's very, very difficult to do on surveys, but you could do in that type of, in that type of in depth method. And that's just a list of publications that came out of my dissertation, partly mentioning that, as you can see, one was published in 08, so you won't be publishing from your dissertation for many years to come because I got my PhD in 03, so just to give you an idea of how some of those things work. But, mainly, I show you this to mention that I used...the dissertation was partly, had substantive questions about inequality, but it was also a methodological project. And partly what I did was, in addition to collecting these in depth data, I also collected survey data and then tried to figure out what survey measures could be good proxies for these actual skill measures because something like this where you meet one on one with people, that's not scalable. All right? You can't do that with a thousand or two thousand people, and you can't...I mean, it's extremely difficult. Okay if you have endless amounts of money, but no one does. Then you could. But since that doesn't exist, you try to come up with methods that are more scalable. So what I did was, come up with survey measures of skill that I could apply to more generalizable [inaudible] populations, examples. So this was a question that I ended up using for skill measure that I have used since and some others have used since as well. It certainly needs updating, but, in fact, it's quite robust over time. And we've had three publications come out that are, those three papers are all methodological. So professionalization lesson number X, think about your projects partly as methodological projects. And especially in this area, a lot of you are probably doing innovative work, methodologically, and think about those as separate parts of your project that you might be able to publish. It's definitely worth pursuing. So one of the things that was helpful...well I guess I haven't told you about the datasets, so never mind that little bit. We'll come back to that. But this is what it looks like on the survey when I'm asking people to rate their level of understanding of different items. As Jenny said, I mainly use surveys in addition to some of those in-person observational projects, and I've actually done several of those over the years as well. So in an ideal world, I would have done a survey on a nationally representative sample, and I actually had a little bit of funding left for my dissertation, so I did a tiny bit of that, but that undertaking is super expensive, and I'm happy to talk about that later as well. And I didn't have funding like that when I was a junior faculty member. It's very hard to get that kind of funding. So instead of the ideal version of sampling, nationally, what I did was, I ended up focusing in a particular population in Chicago, and that was the University of Illinois, Chicago. You may have noticed by now, I'm at Northwestern, so why UIC? Well I have mentioned, I'm interested in questions of social inequality, and as much as we strive for diversity at Northwestern, we don't have quite the diversity of students that would be helpful for some of the questions that I'm

interested in, so I started looking outside of my campus. For those not familiar with Chicago geography, UIC and Northwestern are about twenty miles apart. There are several other schools in the region. That's just actually a small sampling of the schools in the region, so why did I pick UIC? Not for convenience reasons for sure. We collect data in the dead of winter, in Chicago dead of winter, in person, so you can talk to my RAs whether that was a convenient undertaking. No, UIC, there were a few reasons, a few substantively really good reasons to work with UIC, methodologically good reasons. Substantively, it was excellent because UIC is one of the most racial and ethnically diverse universities in the country. So for someone who is interested in racial ethnic diversity, SES, socioeconomic status, that was a fantastic campus to work with. The other thing that worked really well is, UIC has a course that's required of every student on campus to take. So instead of having to go with a biased sample of who took some class that I happened to pick, there's a program that everyone has to take. And they were very kind to work with me on this project. And I should note, because this might seem simple, there were like ninety sections of this course. It was not like we walked into one huge class and surveyed everyone. We walked into almost ninety sections, so a very complex undertaking. I should also note that figuring out, just as you think about your various projects, figuring out the campus and getting the right connections, it sounds pretty easy when I talk about it this way, it took probably a year. So these things take a long time, just to recognize that you don't wake up one day and realize, oh, yeah, there's this perfect campus downtown and they have this great course that everyone has to take. I mean, these things take research and talking to a lot of people and contacting people, etc. And then what we do is, we pretty much every survey I conduct, we use paper and pencil methods. Obviously, using online methods would be much more convenient, why don't I do that?

>> [Inaudible]

>> Eszter Hargittai: Yes. So since I was studying digital equality, it would be problematic to sample on those who are more skilled to take the surveys, have more autonomy, and having time to take the survey, whatnot. So I...it's very important...this is another general methodological research design lesson, make sure that the methods you employ are not associated with the substantive questions of interest in your study. So first, actually, before the 2009 study, which I will be reporting on and some other more recent data, I actually conducted what I considered a pilot in 2006 where I was the one schlepping down to UIC, then I got a grant from the MacArthur Foundation, and I was able to hire other people to go down to UIC, and did a survey in 2007. Using lessons learned from those two projects, I then rebooted and started a whole new cohort in 2009. So that data collection happened with these details. What's really exciting about this dataset is that I have actually been following these people over the years, so the same people, so that's what's called panel data. So a year later, we resurveyed those who agreed to be recontacted, which was a thousand sixty-six people, leaving a little bit more detail about the methods than usual because I'm hoping this will be useful. And a year later, we got a forty-five percent response rate, which was pretty exciting. What's really exciting is that this summer, we're almost done, we've almost closed this, but not quite, we recontacted, again, the initial 2009 people. Why is our response rate higher now? It's because we've actually lost a lot of people through address changes and whatnot. So, well, there are different ways to calculate response rates, and if it's something you're interested in, I recommend AAPOR, the American Association of Public Opinion Research, that has published different ways to calculate response rates. It's a whole other separate field within survey research. Anyway, what's really exciting is that...it says plus because I know that we have a few more surveys coming in, but, basically, we've had over five hundred people this summer take the survey again. So this is super exciting. If you do the math, they were first years in wave one, they were seniors this past year, so now I have such outcomes as who graduated in four years, which is a really important outcome. And I can tell you that less than half have. So that's really interesting, and it can be one of those real-life outcomes that I can start looking at with the data. Because, again, these are the exact same people who took the wave one and wave three. Just one more little note methodologically, and just to clarify, wave two and wave three were both done in postal mail. All right? Because they are not in any one class anymore. Some have left UIC by this year. Many have left UIC, so I'm

contacting them through postal mail and a little bit through email, which looks like this. Send out a letter, introducing them to the project and of the surveys, send out a couple reminders, send out another survey, send out another reminder, and, yes, it's very complex. And then just another quality control that I highly recommend for anyone who uses survey research is, I've had this question on every survey I've done in the last few years, but this time we actually included a second one that's similar, so the issue here is, survey validity, right? So anyone who didn't answer very often to this question, we just pitch, we just don't deal with. And then we have another similar question worded differently, but with the same purpose, later in the survey, and basically...so this year, we're doing a double check, and only people who got both of these questions right remain in our sample. And over the years, it's been pretty consistent that pretty much every year it's been three and a half percent who don't answer the questions correctly. Actually, what's interesting is, this year, even though we had two verification questions, I think we've had a lower rate of missing, of getting it wrong, so I'm really encouraged by that. I'm not quite sure how to explain it, but I'll have to think about that. But I do highly recommend including this kind of a question when you do survey research. Yes, you end up with fewer respondents, but you have less noise in your dataset, which is really important.

>> Do you just hide that in the middle?

>> Eszter Hargittai: Yeah, so this time I was fairly strategic. I had thought about this less earlier. This time, I put this one at the end of the...it was a six-page survey with a ton of questions. I spent many dozens of hours with the layout. So this was at the bottom of the first page, partly to signal to them that I'm paying attention to whether they're paying attention. And then the second one was page four, so over half done, to see if they were still paying attention then. Yes?

>> Does that bias your sample set, the people who read the internet carefully as opposed to those who read it the way most people would...

>> Eszter Hargittai: It might, but the survey data would just be invalid for the other...I mean, if they're not answering the questions I'm asking them, then there's nothing I can do with that information for the survey. We have other studies where we look at people, the same searching studies that I showed you. It might, but there's...it's just for survey research, those are just not valid responses. There's nothing I can do with that. And actually, just to add to that, I've done enough research on credibility assessment, both using surveys and in-person observations, that I'm pretty convinced we simply do not have the right instruments to ask about those things on a survey, so actually don't ask about credibility assessment types of questions. All research projects I've done in the past about that have shown that the survey responses don't correlate in any way to what people actually do for credibility assessment kinds of questions, so I just don't even bother. I mean, I think it's another area where we need methodological innovation because it would be great if we had survey questions, but we don't. I have one type that tries to get at that, having to do with being able to read URLs, but it's just, it's really one small type of approach. Some data about the 2009 original sample, so as I mentioned, it is a very diverse sample when it comes to race and ethnicity, as noted; however, in a couple of variables it's a very homogenous sample, right? So education in 2009, exactly the same, everyone is in the same university, same cohort, and age, they're practically all the same age. This is helpful, however, because we know that both age and education are important predictors of internet use. So by controlling for those variables, we're able to disentangle what else might matter in internet uses. So there can be an advantage to having a sample that's homogenous on certain variables. That said, I want to recognize that it's not a generalizable sample beyond certain variables, right? So what I find, you can't necessarily generalize to eighty year olds. It's very much the wired generation...oh, and, by the way, if you're sitting there thinking, she's going to tell us about 2009 data, that's so old, I will show you, to some extent, that some things are actually quite consistent over time, and also, less interesting, I think, are certain percentages in some cases and more interesting are patterns and relationships of variables, and, again,

those tend to be consistent over time. This is just to show that, yes, it's very much a wired generation, indeed, they're online a lot, they have a lot of experiences. This is worth noting, just to confirm that there are no questions of traditional digital divide, right? Everyone is online. So if we find differences in uses, it's just not because some of them have only been using the internet for two months. That's not an issue. And just to confirm, they do still use email. For years people have been saying that young people don't use email anymore. It's not true. I mean, they might not use it for the same purposes as older people, but for some things, they do use email still. So I showed you what the skill measure was, and this is just some basic information about where people rate themselves in 2009. That paper that I did with, or one of the papers, the methodological paper, I referenced that Patrick and I wrote, most of the data for 2009, 2010, and a couple of other datasets, and one of the things we show in that paper is that people's ratings of their understanding of these terms are remarkably consistent over those two years. And one of the things I'll be looking at with the 2012 dataset is to see how consistent are those skill ratings over this extra year. Yes, Alex?

>> Alex: How often do you refresh these? I guess, [inaudible] delegate figuring out what's relevant nowadays [inaudible] but, for example, frames? I don't know...

>> Eszter Hargittai: Sure. No. Very relevant question indeed. I have kept the same terms for consistency. So one of the issues of longitudinal survey research is to be able to compare over time, you have to have the same exact instrument. So I have kept the same terms, but this year we've added additional ones. So this year we have some that are Twitter specific, and we have some that are privacy specific, and I don't have those data analyzed, but we'll have that, so we can now have...because another issue that comes up that one of you, or probably several of you are thinking, well, one skill, I mean, there are so many domains of skill, even I spoke about different domains of skill, can this one instrument really capture it all? And part of my argument, that it does capture something, but there is the argument also that there are dominantly different domains of skill, and can we come up with proxy measures of the different domains. So that's probably what we're trying to do this time, a social media skill subset, a privacy skill subset. We'll see how that works. Yes?

>> Do you ever map their survey, people's survey responses to their experimental components, like actually, so they say that they could remove a firewall, for example, in the survey, but could they actually do that?

>> Eszter Hargittai: So back at the dissertation work was correlating their information seeking with this instrument, so the instrument came out of how this was...I had all sorts of other survey measures, and this correlated best. So there were things like...there were a bunch of other questions I asked, and I think the 05 paper should talk about some of that. And one of the things I also showed was that proxies like, how long have you been a user, or how much time do you spend on the internet, are not great for skill. I mean, they are related, but this measure is more related. And I mention that because several papers out there talk about skill, and they use those measures for skill, which is a bit of a pet peeve of mine because I feel like, well, if you have frequency of use, call it frequency of use, don't call it skill, or if you have veteran status, call it veteran status, don't call it skill. So that's just another comment on methodology. But, yes, you could definitely do more of that now. The thing is, these are huge projects, right? It's they're...they sound simple, like, oh, yeah, let's see if they correlate, but they take a lot of work. I have a student now who might do some of it as her dissertation, Eden Litt, who was here last year. Okay. But one of the ways that I can do that is, not so much the observations, which is very labor intensive and expensive, one possibility is to ask as a multiple choice question whether they actually know what that terms means. So here you might be surprised, like, how can BCC be so low, I mean, I could people not know what BCC is? And I don't mean, like, do they actually know what the B and the C and the C stand for, but what's the functionality. And so in the 2010, we asked about this on a multiple choice question in terms of functionality, and a third of them didn't

know what it is. So the fact they don't know what it is, which is interesting, and I think notable that, for example, for employers to recognize that their incoming employees might not know what some people might consider pretty basic internet skills. And then there are other terms on there that are more web to point or related. I just grouped all of them together into one index, but one could do different things with them. And basically over time, there are just a couple of these that have really changed, just going back to some extent. I like this question, not in terms of adding measures, but in terms, but, how is this over time. Tagging has become much more understood over the years, I think, with Facebook taking off so much and photo tagging would be one reason, and tab browsing is another one that has...because in these few years, that's really taken off, that browser feature. Yes, Bernie?

>> Bernie: I'm starting to get [inaudible] these or any other sort of, like, you know, is there a niche or a group that know RSS, or is it if you know RSS, you also know social [inaudible]

>> Eszter Hargittai: Oh.

>> Bernie: ...you also know fishing?

>> Eszter Hargittai: Yes, that's a great question. That would be a neat idea. I basically average them. I just create one mean score. And the reasons I do that, and I've purposefully not wanted to read too much into one particular rating because it is marked as an index, but that's a really interesting methodological question that would be cool to pursue. If I had a grad student who is interested, I would love to pursue. Yes?

>> From a methodological standpoint, these are [inaudible] items, like...

>> Eszter Hargittai: It was the item you saw. It was...I mean, it was one through five, and it was presented as one through five, but, I mean, do you want what the labels were, or...I had them on a previous site.

>> One of the comments that I get on the [inaudible] use is that you can't compute means and use any integral statistics with [inaudible] data because it's ordinal.

>> Eszter Hargittai: Okay. Well this is how it's presented. I think it's convincing that this is not ordinal. I don't...I've never gotten that in a review. Okay. Skipping ahead. Okay. So then, as I had mentioned earlier, I mean, that is, it's presented to the respondents as one, two, three, four, five. I was interested in how background variables relate to skill. So one of the findings is that women tend to rate their level of understanding lower than men. This is actually a really complex issue, and because I'm not sure how I will be doing on time and everything, I want to get through, I've removed a few slides, but if people are interested, follow up with me on this because I have a few additional slides, and I think this is a really interesting question and definitely requires more work, I think, the gender aspect in particular. And then we see that racial ethnic background relates to skill as well as does socioeconomic status. Parental education is what I use for as a proxy for socioeconomic status. Overall, if we take gender, race and ethnicity, and parental education together, what we find is that a Hispanic woman in the sample whose parents have less than a high school education has that skill score on a one through five scale, and an Asian male whose parents have a graduate degree has that skill score. So that's a huge difference within the sample of young adults who are all supposed to be super savvy with the internet, right? Okay. So how about nationally? What can we say nationally? Well what's nice is that... so one of the things that I encourage you to do is what you're doing here, get to know people, other people interested in your work, interested in things that you're interested in, but also agencies and people who work at places where they might collect data. So fortunately...so the question was, okay, we find these differences within this one population, how might that look nationally with a more diverse sample? And fortunately, John Horgan, who was, at the time, at the

Federal Communications Commission, put a subset of my skill item on a national sample survey in 2009, so that was great because I got to look at data, and those data were made public. They are public. So I was able to look at what predicts skill, what relates to skill in a more diverse sample. And that sample had information on income, also had diversity in education and age, so I was able to look at those variables that I couldn't look at in my sample. You can ask me later why I didn't have income data for eighteen, nineteen year olds, but basically you can't really ask those age groups what their income is. It's a completely useless question. We can get into it later why. But...so based on the national sample, what we found is that income is positively related to skill as is education, so concerns about social reproduction here potentially, but what's also interesting is, and I haven't framed today's talk as a talk challenging digital native discussions, which is the assumption that everyone young is super savvy with the internet, but a lot of my work challenges that assumption. And what I found was that if you look at people fifty and under in the sample, there's actually no relationship between age and skill. So I...if you do work with datasets of people with diverse ages, I encourage you not necessarily just to put in age as just this one continuous variable, which, I mean, of course that's helpful to do in some ways, but do some more basic parceling of the data to see whether it's really just like people at the high end of the age distribution who are skewing the findings, for example. Because, yes, over fifty, there is a relationship between skill and age. But most of the people who talk about digital natives are not talking about forty-five year olds, right? So this is just important to recognize. Okay. Another question, is adoption of a service random? And this is where I especially relate this to methodological and research design issues. So we already seen talks here, and I suspect you'll see other talks, about all the opportunities of grabbing data, log data, automatically generated data from various websites, and using those as your core dataset, and one of the challenges there is, well, it looks like, who adopts different services is not randomly distributed, not shockingly, I think. So, for example, what I found, and this is based on 2007 data, so that initial dataset I mentioned was that race and ethnicity related to Facebook and MySpace adoption. And this is where you may have already seen, just quickly going through this, what I also found was that this was persistent over time, and this is where I come back to that issue of the numbers might change over time, but the relationships of variables don't change that quickly, per se. So even two years later, we still had that relationship, especially to MySpace use and adoption. And then, one more thing, here, just more of a professionalization note, but also showing the strength of the UIC dataset, so in 2007, when I published the paper about the racial differences, the Chronicle of Higher Education wrote a paper about it, but when I tried to get more nationally, more national media to cover it, the response I got was, oh, who cares about a bunch of Chicago young adults. What was interesting though was that two years later, Nielson, based on a nationally representative sample, came up with pretty much the same result, and then the press corps covered that broadly and widely, but I basically already found that two years earlier based on my dataset. So it just shows that if you do careful sampling and really think carefully about your research design, you can actually get really great data, even if they're not nationally representative, per se. So in 09, four percent of the sample used Twitter. Just to anchor you, because it's hard to forget these details, early 2010 was when Oprah joined Twitter. There was just a lot of buzz in early 2010 about Twitter. Already in 09, but it was really about 2010 everyone was supposed to be using it. In this young-adult sample, eighteen percent were using it. That's less than a fifth of them. Then I was interested, okay, well who uses Twitter? Can we predict that? Can we explain that? So one of the things we found was that blacks were likely to use Twitter than others, which was consistent with PU data and other studies, so that was nice that, again, we had similar findings. But what no one had really been really able to explain was why it was that African Americans were more likely to use Twitter. And there were some hypotheses around, but no one had the data really to look at that. And that is where longitudinal data are really valuable. So what we can do here is that we can actually take data from a year prior to adoption and see how factors, features of people from a year before they adopted Twitter influenced their adoption of Twitter a year later. All right? So, for example, there might be some factor that if you had just cross-sectional data, data from a point in time, and you found a relationship, you couldn't say if it was that factor that influenced Twitter or if it was Twitter use that ended up influencing that factor. But when you have this lag in the data, you can say, well, no, that happened prior, before they used Twitter, so then you can isolate an effect. So not

surprisingly, based on what I've shown you...oh, and what I'm doing on this slide is basically showing you graphically the results of regression analyses, and in the paper that I cite, or that I cited, Hargittai and Litt 2011, you can see the regression of results in detail with the coefficients. So two hypotheses people had brought forward as to why African Americans use Twitter was that they text more and that they are more likely to have more access on their mobile devices, both of which are true. Both of which are factors we had data on for 2009 for these people. We tested them. Those things didn't seem to matter. What did matter was internet skill. People who were more skilled in 09 were more likely to be Twitter users in 2010 and also, and I'll do this one more time, watch the line that disappeared, so we found that people with interest in entertainment and celebrity news were more likely, so a year prior, were more likely to use Twitter in 2010. More importantly, that pretty much explained the racial difference. So what's really going on, for those of you who don't quite know statistics, is that it's really that African Americans in the sample, on average are more interested in entertainment and celebrity news than whites. But if you take a white person and African American in the sample, who were equally interested in entertainment and celebrity news, they're equally likely to have adopted Twitter. So that was really what was driving it, not race, per se. I will also note that we looked at interest in some other topics that have been associated with Twitter use potentially, and interest in news and politics was not related to adopting Twitter. This doesn't mean the use of Twitter might not increase people's interest in politics and news, but it means that those were not interest areas, at least for this age group that drove their adoption of Twitter. And a year ago, this is what Twitter homepage looked like, which was interesting, because it showed that they had figured out that interest was what was driving people to Twitter. My idea was from 2010 to look at this, so I was feeling pretty proud of myself. Okay. Moving on. Participation gap. Let's see. I'm just thinking about time because there are a few more points I want to make. As I mentioned, one of the things I'm interested in is who...Jenny, how are we doing on time? I'm not sure how much time we want to leave for...

>> Jenny: You have twenty-five minutes;

>> Eszter Hargittai: I do? Okay. I just want to make sure we leave time for Q and A. So I'm going to take two minutes to do audience participation on this slide because I think the points come across better if I have you think about it just briefly. So these are six activities about which we asked on the survey that whether people had engaged in them, and we actually asked how often, but the frequency was not huge, so I'm just showing you ever, whether they had ever done these things. But note that I'm not listing these in order of popularity. I am, however, listing them in some order, and I ask you to work with me on figuring on what that order might be. So what logic to this ordering can you think of if you haven't seen me give this, give a talk about this material before? Yes?

>> It seems more...was there social or friend participation and you encouraged more, presenting like more formal or [inaudible]

>> Eszter Hargittai: Hm mm. Yeah, that's great. And just to clarify, the creative quiz was probably interpreted at that time by respondents as like a Facebook quiz. I wouldn't read too much more into that, but, yeah, that's a good point. Other thoughts on what the ordering may be? Yes?

>> Internet activity, the first one, the top, you're taking input from other people...

>> Eszter Hargittai: Hm mm.

>> The last one, just throwing it out there.

>> Eszter Hargittai: Okay. Great. Yes?



>> Ben: Men and women.

>> Eszter Hargittai: But you saw this last year.

>> Ben: No, no.

>> Eszter Hargittai: You didn't?

>> Ben: I don't remember that.

>> Eszter Hargittai: You didn't? Okay.

>> Ben: Now listen, you're below the portion of women participating Wikipedia in the high proportion of women in your sample.

>> Eszter Hargittai: Okay. So Ben has pretty much nailed it, but the other two comments are, I think, very important to remember because I think they may be explaining partly the gender difference. So basically, I've listed this by variation and participation by gender. So on top are activities where there is less difference between women and men in the sample, and on the bottom where there's a huge difference. Right? So this is huge that almost a third of the men say that they've edited Wikipedia and less than ten percent of the women say they've done that. So think about the two comments we had because I think that's really interesting to map onto this, right? So whether it's the, you're socially getting information versus you're putting your opinion out there to the rest of the world. Your point was something similar, or feel free to repeat it. I'm just...

>> No, I think the social interactive versus...

>> Eszter Hargittai: Right.

>> Expressing an opinion of authority.

>> Eszter Hargittai: Yes, exactly. So that too, I mean, I think may well be explaining the gender variation of, you know, the authoritative voice that you're putting out there, which would relate to other types of usage. Yes?

>> What about like creative versus informative? Because wasn't the top seem a little more like generative whereas the ones on the bottom seem to be more information...

>> Eszter Hargittai: It could be. Yeah, I mean, there...so this is one of those areas where I think there's room for tons more research, right? Partly, the question of who participates, but also, why do we find differences among different types of people in terms of participation, part of which is very much gender related, right? So Wikipedia, itself, has put out information about just what a tiny proportion of its contributors are women, for example. The one thing where women did more was change the privacy settings of their Facebook accounts, not too shockingly, and actually by 2010 that was almost universally, just almost everyone had done it, so there wasn't that much generation. So then I was interested in, well, how does people's background relate to how much they participate in these activities. And first, reality check, a quarter of the people hadn't done any of this, so, again, not the super savvy digital natives, per se. Based on what I just showed you, this isn't surprising. So women reported participating in few activities. There was a relationship with race and ethnicity as well as socioeconomic status, so overall what we found, what I found was that if you take the average Hispanic female whose parents have less than a high school degree, she

has done one of the five things. And if you take the average white male whose parents have a graduate degree, he has done three. That's a huge difference on a zero to five scale, huge. And then what I also found was skill relates to participation. Of course, here, you know, the more you participate, the more skilled you are, although this is again why we can look at the data over time and show that prior skill does relate to later participation as well. So now just to wrap up, I wanted to talk a little bit more explicitly of why you should care about this even if you don't really care about digital inequality, per se, but have other questions of interest. So one of the...okay. Does anyone know who that is?

>> Hm mm.

>> Yep.

>> Maybe.

>> Eszter Hargittai: Yes, Larry Page, one of the cofounders of Google. So when I was participating in a Google faculty summit, and he was going around asking everyone questions about their work, and when I told him that I was doing information seeking types of things, and this was more the observational work, he said, well, why would you need to do that, we already have lots of data on how people search. So the question is, well, is it enough to look at log data, or can we get information from other data sources? And my argument...well, first to clarify, it's log data, but I think in this group, probably most know what it is, but it's basically automatically generated data about people's uses of various sites and services. My argument is that, yes, while log data can be really helpful for certain things, other things you don't necessarily get through log data, and that's why other types of data collection might also be helpful. And so part of the issue is that you only have log data on your users, right? A company only has data on its users. Well, but I just showed you that adoption of a site is not random. [Inaudible] means that a company is only going to have data on a nonrandom sample of the population, which could be helpful to answer all sorts of questions, especially the questions that are solely about its users. But if it's a social scientist that you're trying to answer questions about more general societal patterns, then it's probably tricky to rely on a nonrandom population. So as I've said, I've already shown you that adoption of a service is not random. Another issue we know, and here is just one reference about this, is that people understand and use sites and services differently, right? So I might use Facebook for certain purposes in my life, and you might use it for different purposes, so is it legit to compare us with respect to just one specific social science question? Maybe not. And then the other issue is that, what you do online is just one place where you do certain activities, right? So you might communicate with your friends on Facebook, again, as an example, but you probably do so using other media as well, so can a researcher really just figure out your relationship to your family through Facebook data? Probably not because they are lacking data from all sorts of other platforms and face-to-face interactions. Going back to one of the examples I started with, so how does my internet use relate to political participation? Well let's say you have some data and you find that over time more internet use leads to more political participation, great. But based on some of the research I've done and related to some of what I've shown you, it may be that in fact depending on where you got your data, your dataset is actually biasing towards people who are from a more privileged position, and if you happen to have data from a more representative sample, you might in fact find that that relationship only holds for people in a more privileged position, and those from less privileged positions, we see no relationship between these things, or let's be a little bit more optimistic, we do see a lower curve. But in reality, so even if participation would be going up for the less privileged as well, over time, the gap between the less and the more privileged is actually increasing. So depending on how you look at the data, you could have an optimistic or a pessimistic view of what's actually going on, which is partly also just a call for being careful about statistics and recognizing that they can be interpreted in very different ways. And then, you know, if you don't care about political participation, plug in your research question of interest, and the point is that this applies to different domains. And usually often when data get grabbed from online sources, you end up

focusing on the people who are in more privileged positions because they're more likely to be taking part in that activity you're interested in in the first place, which is why I go back to that initial comment I made that you want to use methods in your studies that are removed from the substantive questions of interest in your study. So you end up having less data on the less involved and pretty much no data on those who don't do that thing at all. And so this is just a challenge to those who use these various sites and services of data sources, and I've seen studies used pretty much all of these as source of data, but, again, recognize that those who don't use these sites are down there at the zero, right? So they're not in your dataset at all. And what does that imply? So what that implies, just to wrap up, is that data from log sources are rarely, if ever, representative of a larger population, so just be very conscious of that, and be careful. This doesn't mean...I mean, just to be clear, to get super excited when the kind of opportunities log data provide, we just need to be very careful about the scope of the questions we're asking and the generalizations we're making when we use such data. The other thing, just back to skill, different papers I've worked on have shown that skill relates to very different types of outcomes in terms of both information consumption and production. So we need to think about skill. Why is it helpful to think about skill? Well other variables that matter like socioeconomic status are not things that we can just willingly change, and I'm not saying skill is something that you can change easily, but I think there's room for interventions and seeing how improving people's skills might lead them to use the internet in more diverse ways and more participatory ways. So that seems like a much more realistic lookout for intervention than some other things that we know predict internet uses. And then implications of this for society at large for those who are interested in questions of digital inequality, there are different possible outcomes, right? So one possible outcome may be that people from less and more privileged backgrounds equally benefit from their digital media uses, so not much changes, difference between them remains the same. The optimistic view is that people in less privileged positions benefit more than those in more privileged positions and so the gap decreases. I've shown you pretty much no data to suggest that this is what's happening. The alternative is that those from less privileged positions are doing fewer things online and thus benefiting less from their online uses and thus the gaps may increase. There's a bias in this entire slide, which is everything is going up. So there is also the alternative, which is that the slope actually goes down and that it might go down for those who are less privileged, increasing the gaps even more. So to conclude just some of the points that I've tried to convey to you. Adoption of a service is not random, so think about the platform if you're basing your research design by grabbing data from one platform. Think about how the users of that platform might be biased in one way or another. We know that skill relates to internet uses and those who are more privileged tend to be more skilled, so those who are more privileged tend to benefit, tend to be engaged in online activities more for which they might benefit and increase their life chances. In 2006, you might recall, Time came up with its person of the year as you, but what I've tried to convey to you is that really that's just some of your, more precisely some of us in this room, so just keep that in mind as you do research in various populations. And I just want to thank my funders and people in my lab, who have supported, who have assisted in a lot of this work, and open up for questions.

>> Great. [Applause]

>> So I guess my question, you chose these particular terms, and I'm interested in like how you got to the place in choosing these terms because there's nothing [inaudible] social bookmarking and RSS feeds, and I think some of those are very specific, and I guess I feel like there's not necessarily something missing, but the next step is, how do you [inaudible] question is, do you have the skills to decide how much information you want each day and manage that information and remember that information for later use, and so is it even directed to people who know those particular terms and missing out on people who might have alternative strategies? Because I feel like a lot of time in internet research we get into the, this is the most strategy, so we're going to look at people who use these strategies, and sort of how did you get to choosing that as the best set of skill or terms to ask about skills?

>> Eszter Hargittai: Hm mm. So, again, I refer you to those papers from that discusses methodologically in detail. It's partly related to what was around back eight, nine years ago, a decade ago, but also I encourage thinking about that as an index, again, made up of twenty-seven items. I don't ever look at any one of them alone, so the idea is, how does it all average out, and it seems to be related to a lot of things. That said, we are updating the instruments, so that's an issue. I do think that over time, as the internet has evolved, as I mentioned, different domains of skill come up, and I think it is important to pursue research on different domains of skill. So I very much encourage people in this room to be doing work on that, and the topic you mentioned could very much be like an information management skill type of concept, and I would love to see work on that. And then I would encourage people to try to come up with proxies of measures that they developed if they do in depth studies of something like that, which I think is usually necessary to do something in more in depth at first or more qualitative, but to do it in a way that we can then have measures that can be put on instruments that can be applied to more generalizable samples. Yes?

>> So the sample that you researched had, there was no difference in the access to internet, ninety-eight percent of the people had internet access at home.

>> Eszter Hargittai: Hm mm.

>> So what are your thoughts on doing similar research in populations where there is different [inaudible] access to the internet [inaudible]

>> Eszter Hargittai: Right.

>> [Inaudible]

>> Eszter Hargittai: Right. Hm mm. Yeah. So that, I think that's an important variable generally in terms of where people use the internet, and it turns out that actually autonomy of use, which is a concept that's come up in the literature, which I've operationalized as number of different locations where you potentially have access, which is really what you're talking about, right? Different locations?

>> Hm mm.

>> Eszter Hargittai: It turns out it's an important predictor of a lot of what I looked at, even in this population. So I'm sure that it...I mean, I would suspect that it would matter quite a bit elsewhere. What I would do is, collect data on the location of where people have access and make that part of the analysis, and I still do that, even with this population. And, again, one of the most consistently important predictors of different usage has been autonomy, so I suspect it would matter. I would just keep collecting nuance data on that. Like I list ten different possible access locations for people to check off.

>> And that would also relate to the socioeconomic status...

>> Eszter Hargittai: Yes, absolutely. Exactly.

>> [Inaudible] to access...

>> Eszter Hargittai: Yep. Yes. Very much.

>> [Inaudible]

>> Eszter Hargittai: Oh, I'm sorry. Yeah, I know. I saw her, yeah.

>> So what is it about gender? Do you take like an evolutionary perspective, or is it that, you know...

>> Eszter Hargittai: I went to a women's college, I don't take an evolutionary perspective. [Laughter] No. So there...I mean, I don't think at a deeper level there are a lot of things going on, but...so let me just quickly show you this, and I refer you to these papers that I think were fascinating findings, although I'm obviously biased. But what we found from data from my dissertation was that women actually did not differ from men in their actual abilities to find information that much. It was their self-rating of their skill that differed. Even when you held for actual skill, there was a gender difference in self-rating. Now the self-rating we used in this paper was just simply asking people, what do you think your skill is, which I discourage from doing because research, my research suggested that's more gendered than asking people to rate their skill. But even in the rating there seems to be a difference. But let me...so first of all, I think this is partly what's going on. I think this is a great cartoon, and I think, so no, I don't think it's evolutionary or whatever you called it. I think it's socialization very much, and you need to spend some time on certain forums to see that perpetuated all the time online, and I suspect also still in physical environments. But the other thing is, we had a paper where we looked at who contributed online, different types of content that I talked about today, where it looked like men were contributing more. But then when we controlled for skill, that went away. So even if it's only in people's heads that they are less skilled, it seems to influence what they do. So that's why it's important. That's why socialization, I think, matters so much is that even if it's just that you think that you aren't as good, you're going to be doing less. So, no, I think it's socialization.

>> Would you be interested in looking at like younger populations to see, you know, before they're really social...

>> Eszter Hargittai: Yeah, I mean, I think that would be great. It would be really important to look at younger populations. That's not my focus, but I definitely encourage people doing that.

>> Ben: I just...inspired by John Robinson, he preferred to do actual performance than people's impressions, so he would ask something like, when was the last time you used X?

>> Eszter Hargittai: Hm mm.

>> Ben: Or how often have you done? Not just, I know this...

>> Eszter Hargittai: That is my measurer.

>> Ben: I thought you were saying how fully you know it, the one through five or how fully...

>> Eszter Hargittai: That is, but the outcome here is how much have you posted online and how much have you actually shared, x, y, z. That's the outcome and the predictor is your level of understanding. No, absolutely, I do the actual experience as the participatory measurer.

>> Ben: Then you should know my enthusiasm for your work ten years ago remains strong, but my enthusiasm is not pervasive. Your work was unique and distinctive at the time, and, you know, I was promoting it, and there you provided the data that was really there, that, you know, these things happened.

>> Eszter Hargittai: Thanks, Ben.

>> Brian...

>> Eszter Hargittai: She's had her hand up for a while. Yeah, go ahead.

>> So this is about your income question. So, I'm just wondering, this may be because I come from Drexel, which has a really big co-op part of the environment, so work while the students are at school is very important to them. So I'm just wondering if you asked about like, if they were part of a work study, and if so, what kind of work study, are the work studies likely to be more or less technical?

>> Eszter Hargittai: Hm mm. Sure. Yeah. So, first of all, I don't have an income measurer. The one income measurer, that was in a national sample of people of all ages.

>> That's what I'm wondering...

>> Eszter Hargittai: Right. So the reason I don't have an income measurer is because half the people at UIC live at home, so they're just...and people don't know their parents' income, and there are just a million reasons that...and income of a nineteen year old, none of them are full-time work...or very few work full-time, etc. It's just very hard to compare. That said, I do have data on whether they're working, how much. Those have never been important predictors. I also have data with various technical experiences, like have you ever taken a class that was related to tech stuff, never has been relevant. So I have looked at various other factors that I really didn't talk about today. They don't seem to matter that much. I suspect a place like Drexel, it's tricky because they're probably biased towards a lot of more info savvy types of people, so that would be it. It would be an interesting sample to look at, but one would have to recognize that.

>> Okay. I said, Brian, but I think I met Bernie.

>> Eszter Hargittai: Bernie.

>> Bernie: Well I'll try to make it short because we're running low on time. I do want to make one important comment though for the students. You'll notice in Eszter's presentation that she did things like, you know a person of this age and so forth, these are margins. These are what you do when you digest your data. You don't just present a regression table. Digest that data and actually start presenting the results from that. It's really illustrative and communicates this very effectively, and so kudos, but, yes, don't just present a table. Digest data. The second, what the question is, what about mixed signals from people who are very [inaudible] decided, like I don't need Facebook or I don't want this or I'm dropping out of this and have variegated uses of these tools, and so it's not more tools equals more skill, select tools equals more skill?

>> Eszter Hargittai: Yeah, I think that's fair, and I'd like to think that some of the papers I've worked on are a little more nuanced in that way, or like Patrick and I have a paper where we came up with the topology of social networks site usage where it's not necessarily more, but it's more like how diverse and how frequent, it's more of a two-by-two of that. And I absolutely appreciate that point, and I think, I need to try to get at that in different ways in the future. Yeah, I would love to talk to you about that. I have various thoughts. But the data point Bernie made, another point that I actually forgot kind of in the beginning and went back to, there is so much you can get from your data just from crosstabs. I don't think that can be an end in itself. Like you need to keep going from that, but just by various statistics and figures can be super communicative. So in some of my papers, I kind of forgot that, and then I went back and realized, well, that's like, that really conveys information also to people who don't know that much stats, so I do encourage looking at your data at various levels.

>> Okay.

>> Eszter Hargittai: Okay.

>> Thank you very much.

>> Eszter Hargittai: Thank you.

[ Applause ]