The digital production gap in Great Britain: How sampling, mechanisms and theory matter with digital inequality?

Jen Schradie

The digital production gap in Great Britain

How sampling, mechanisms and theory matter with digital inequality?

Jen Schradie, Department of Sociology at the University of California, Berkeley

Published online: 06 Jun 2013

Jen Schradie

THE DIGITAL PRODUCTION GAP IN GREAT BRITAIN

How sampling, mechanisms and theory matter with digital inequality?

In the iCS article, 'Who Creates Content?' (April 2013), Blank contends that digital production inequality depends on the type of online content in question. Using 2010 survey data from the UK, he uses a principle component analysis to cluster content activities into three types and conducts a logistic regression analysis with an eye toward social class. He finds some levels of inequality but also finds parity with two of his content categories. These findings differ from a substantial body of literature from the United States. Rather than explaining these differences between Britain and the United States with substantive and theoretical reasons, Blank attempts to find fault with the methodology of this existing literature. In this comment, the author shows how Blank's analysis does not reveal as much digital production equality as he claims it does because of his misinterpretation of causal paths of inequality, as well as problems with his operationalization of online content, which conflates Internet activities and online content. In the process, the author explains why Blank's resulting critiques are misguided. Finally, to explain better his unique findings and to help advance the field of digital inequality, the author suggests discrepancies between Grant's findings and previous research may be attributable to differing study populations: Internet users versus the general population; the age differences of respondents; the timing of the studies; and between-country variations. At stake in this debate is the reproduction of social class stratification with digital technology and content creation.

Routledge
Taylor & Francis Group

Introduction

Grant Blank's article in the CITASA special issue of *iCS*, 'Who Creates Content?' (2013) uses a survey of Internet users in Britain, where digital content production has not been thoroughly studied. The paper raises a critical question missing from existing digital inequality studies: how does social status map onto the *type* of content being produced online, such as political or social topics? Additionally, Blank's analysis is unique in studying political content and its relationship to social class.

In line with my findings of a digital production gap, Blank reports that people with more education are more likely to produce political content in the form of social media posts and e-mail. This coincides with the results in my 2011 *Poetics* article, 'The digital production gap: the digital divide and Web 2.0 collide' (Schradie 2011). There, I raised the question of how social class maps onto ideas of digital democracy by interrogating who is actually producing online content. I found that, among the general American adult population, as well as among Internet users, educated Americans are more likely to create digital content than less educated Americans.

Blank's other results, however, represent a break from American digital inequality literature, including Correa's (2010), Hargittai and Walejko's (2008)[1] and my (Schradie 2011, 2012) studies of digital content creation. He contends that social class, via education and income, does not matter among Internet-using British who post what he calls 'skilled content'. One of his critical findings even suggests that people from the UK with *lower* incomes are more likely to post online 'social and entertainment content'. He argues that stratification depends on the type of content.

Instead of engaging substantive and theoretical explanations for these differences between his findings and scholarship which finds digital production inequality in the United States, Blank insists that the existing literature is empirically faulty. This is problematic for three reasons: (1) his criticisms do not hold up to statistical scrutiny (2) his own analysis does not reveal the level of equality that he claims and (3) he generally overlooks theoretical explanations and hypotheses that can help understand differences in findings and thus help move the field further along.

Certainly, research on the nature and scope of digital inequality needs expansion. Blank's analysis does move digital inequality research forward, in that he brings another country into the conversation. But his unexpected results do not undermine existing sociological theories of stratified digital engagement practices (Hargittai 2008; Robinson 2009; Schradie 2011) to the extent he suggests.

In this comment, I discuss why we should approach Blank's findings with caution both from a methodological standpoint and a theoretical one. I first

show how Blank's analysis does not reveal digital production equality as he claims it does because of his misinterpretation of causal paths of inequality and due to problems with his operationalization of online content, which conflates Internet activities with online content. In the process, I explain why his resulting critiques are misguided. Finally, I suggest his anomalous findings may be attributable to differences in the populations studied: Internet users versus the general population; the age differences of respondents; the timing of the studies; and between-country variations.

Weaknesses in critiques of inequality and claims of equality

Path analysis as mechanisms of inequality

Blank reports that he does not find typical class inequality with two of his three categories of content creation activities. However, his own results show otherwise for his 'social and entertainment content' category. This misinterpretation demonstrates how the article could benefit from an understanding of the *mechanisms* of inequality via path analysis. For this category, Blank contends that non-elites are producing these types of digital content. However, in his demographic model presented in Table 3, people with a university degree are more likely to create this type of content than those without a degree. It is only when he adds variables that are often a proxy for class, such as the number of devices, technical ability, or Web confidence, do these factors take away the significance of education (and possibly 'give it' to income). In other words, these variables can be class-based mechanisms for digital content production. Therefore, the causal chain is most likely education → Internet skills/confidence → digital content production, which explains why education was no longer significant in this model. Thus, it is imprecise to say that social class does not matter for this category of content creation. In other words, there is no education lagged effect when controlling for these mechanisms.

It is this type of misunderstanding of the path analysis process that explains an unusual footnote (9) in the article in which Blank claims that I admitted an error of statistical significance in one of my Tables in an e-mail exchange. Unfortunately, Blank misunderstood our communication, as I never wrote that. Rather, in our exchange, I explained to Blank how this path analysis works: how the process of adding additional interactions, hence variables, relate to class variables, which accounts for the differentiation in significance between two of my models. For clarification on this process, it would help Blank's article to incorporate the methods of Alwin and Hauser (1975), who describe path analysis processes in terms of total effects, in this case, of education (both indirect and direct effects).

In my article, I both identified and measured intervening mechanisms to the point of having very little *direct* effect of education in some of my models. The original total effect in earlier models ends up as indirect in my final model, none of which would change the predicted probability of education itself. Blank misinterpreted these findings as 'ambiguous' and as a spurious correlation, but this would only be the case if some other variable, such as race or gender, has an equal (or greater) level of exogeneity, *and* if that exogenous variable reduces the total effect of education in kind. This is not the case. In essence, then, none of this refutes the consistent finding of class-based inequality in my study.

Blank also dismisses the findings of other digital inequality scholars to justify his own findings. For instance, he critiques H/W for not having the data to back up their conclusion about educational divides with online creative content. Blank references their Tables 6 and 7, but at this point in their path analysis, they are only including people who are already creating online/offline content, so it is no surprise to find more equality within the population of elite creators. It certainly does not prove H/W wrong.[2]

Consistent and persistent digital production inequality

Blank (2013) also contends that my use of 17 surveys over a nine-year period, in which I find consistent and persistent inequality, results from a methodological artifact: 'Schradie simply assumes without testing that the within-survey and between-survey coefficients do not differ in sign, size, and significance ... Without a multilevel model, her results may reflect differences across surveys rather than differences between respondents' (p. 17). However, this assertion is without merit for two reasons.

First, Blank's imaginative figure (2) is an illustration of the classic 'Simpson's Paradox', which is easily handled by checking for change over time in the associations between Xs and Ys, so a hierarchical linear model (HLM) is not required. I included dummies for each year and efficiently estimated what would be fixed-time effects in the HLM. I have already tested for slope difference by year, so by checking for these time interactions, I have accomplished all that an HLM could. My results would be no different if I put it in the HLM. Simply, I already included a time/year measure in my analysis.[3]

As I pointed out in my article, 'Tracking productive online activities over a nine-year period enables a replication that allows for robust patterns rather than an idiosyncratic spike during one survey time period'. In other words, one survey during one time period also has drawbacks, which Blank should interrogate since his findings are so different from other studies. Again, rather than my work being 'ambiguous', as Blank claims, my findings of inequality are robust, and are substantiated in another paper (Schradie 2012), which reports similar findings.

At the core, then, of Blank's critiques of H/W, Correa and my work, is a futile search for methodological flaws in others' work, rather than interrogating substantive reasons for differences.

Challenges with the conceptualization and operationalization of online content

Blank critiques previous studies for not distinguishing among different types of content. According to Blank, other scholars either thought content creation was all the same (Correa and H/W) or all different (Schradie). Aside from over-stating his own innovation (H/W, for instance, *did* examine different types of creative content), Blank fails to incorporate the changing dynamics of Internet practices and activities, which result in different types of content creation. For instance, the online tools and activities included in my analysis from 2000 to 2008 were much more distinct and predated the blurring of digital practices of today's social media sites, such as Facebook or Twitter.

Blank, tries to organize different activities into content categories using a principle component analysis (PCA). This technique uses a correlation structure to see which online content production activities cluster together (see Table 1).

TABLE 1 Blank's PCA-devised categories of online production activities.[a]

Label	Definition	Online activity from survey
Skilled content	'Require a certain level of technical skill, rhetorical skills, and considerable personal commitment'	• Maintain a personal website • Write a blog • Post writing, stories, poetry or other creative work
Social and entertainment content	'Share a common theme of social ties and entertainment'	• Use SNS[b] • Post photos • Upload video or music
Political content	By what respondents say they do	• Send an email or message supporting a political or social cause • Comment on a political or social issue in a blog, Tweet, or on a SNS

[a]Definitions derived directly from Blank's article.
[b]SNS is a social networking site.

The rapid technological developments in the Internet itself and certain aspects of his methodology, however, undermine his claim to novelty. First, Blank's study does not really measure content, per se, but *activities*. With all of Blank's build-up, I greatly anticipated his analysis of online *content* variation, from cat videos to conservative blogs. Instead, he used responses from the Oxford Internet Institute's survey that were similar to the online production activities that I examined. His method generally uses variables of online activities, such as blogging, being on Facebook or posting videos, rather than the specific content created with those tools (with the exception of questions he uses for the posting of poetry, creative work and political content). His PCA results in three categories: 'Skilled Content',[4] 'Social and Entertainment Content' and 'Political Content' (see Table 1).

Since the dawn of the Internet, we have seen blurring of different Internet activities with the content these practices have produced. While Blank marginally acknowledges this blurring, he still contends that the PCA shows 'most of the time that they spend [*sic*] on SNSs, blogs or personal websites is not on politically related topics' (p. 10). This conclusion is unwarranted, however, for his finding is most likely a function of the lower level of political content being posted in general. Certainly in 2010, blogs and Facebook included political content, yet he separates these out into three variables.

Next, Blank's PCA mixes apples and oranges, as the categories derived from the PCA do not articulate substantively or theoretically. Blank includes activities (such as posting political content) that are embedded into other activities (using a social networking site), yet he counts them as different variables. In turn, the audience for e-mailing political content, one of Blank's political content variables, is vastly different from the others: an e-mail message is generally to just one person or perhaps a few people, whereas all of his other categories are either overtly public (e.g. websites/blogs) or semi-public (e.g. SNSs). This conflation, of putting overlapping variables into the PCA pot, is problematic since this technique is highly sensitive to the particular combinations of items that get entered into the analysis. Thus, we can only conclude that Blank's set of activity categories are clustered, not that they are actually correlated based on content. These weaknesses should have been explored both empirically and theoretically to see if they would change the PCA results.

So rather than this being 'evidence-based', Blank's PCA is a tentative approach lacking a conceptual framework. A PCA works best when we do not know how to cluster our data. Missing is a strong theoretical, normative or practical explanation of his categories. PCA infers items and item domains from common variance and covariance. But Blank's analysis is completely silent on the source of whatever common variance it uncovers. In his case, the exploratory methods fail to exploit the content embedded in his measures. A confirmatory factor analysis would have been useful.

Theoretical differences as explanations of (in)equality – sampling more broadly

Let us assume for a moment, though, that Blank's analysis holds up to methodological scrutiny. What then might explain different findings of digital production from other scholars? Sampling differences between the two cases, I argue, are the most likely culprit.

First, it is possible that Blank found lower levels of inequality because he does not analyze people without Internet access, usually those at the lowest rung of the socioeconomic ladder. My studies of digital production inequality (2011 and 2012) examine two samples. One is similar to Blank's: the adult population of Internet users. I found that people online with a college degree were more likely to produce online content in 6 out of 10 different production activities. That is the analysis to which Blank compares his findings. I also examine online content creation based on a sample of the *entire* Adult population. With this sample, I found that there was a digital production gap in all 10 online tools I examined. In other words, by excluding people who are not online, it is impossible for any scholar to claim that online content inequality does not exist. His argument, which holds that minimal inequality exists, would be strengthened if he had included the general population, since a full $\frac{1}{4}$ of the British population is not online, according to his own data. What are the demographics of the people who are not online? What happens when you include both? How would these same models look if the author includes the entire sample, not simply those online? This is an essential analysis missing from the paper and makes claims like 'social status is irrelevant' overblown rather than definitive. Status matters very much.

Another sampling explanation could be that Blank's sample includes British 'adults' who are as young as 14, rather than the population age in the studies to which he compares, which start at 18 years old. The younger age range might have a downward bias in the education slope. Quite simply, educational attainment is a poor social class measure of those under 18 years old. In an analysis of adults over 18, such as in my study, many of whom have left school, education reflects underlying abilities and exposure to educational opportunity. Conducting this study with adults age 18 and over would be a more apt comparison to my study.

While Blank does have a measure of underlying ability in terms of Internet skills, which education level usually measures, he does not parse them out with age levels. People in school have more opportunities for engagement with digital technology because of access they have at and through their school-based relationships. While Blank says that he isolated students into his PCA and came up with the same categories, it is unclear how social class maps onto the final logistic regression analysis.

Another possible explanation for the differences in findings between Blank's and other studies is the timing. Blank's sample is from 2010 data while mine draws from 2000 to 2008, and others' are from earlier, as well. It is possible that earlier on in the social media explosion, fewer people were digital producers. As a result, the people in his survey may be farther along on the adoption curve of digital content production activities. Diffusion may be part of the answer as to why he finds less inequality. While this may be possible, my study (Schradie 2012) of blogging over this same time period found consistent and persistent digital production inequality based on social class. Nonetheless, this complexity of changes over time requires examination of if and how digital production inequality reinforces or challenges class differences.

The fourth and final sampling hypothesis that could explain some of the digital inequality differences is that the American case is different from that of Britain for broader social, economic or cultural reasons. While Blank gives a slight nod to the fact that educational differences between the two countries could explain the variation, this explanation defies logic. He says that perhaps it is due to the US and British education systems. If this were the case, however, there would be *more*, not less inequality in Britain. While both societies are stratified based on class, Britain has a more stratified educational system (e.g. Kerckhoff 2001), which may extend to educational support, or the lack thereof, for digital engagement. It is much more reasonable to think that there is something else happening that makes Internet engagement more stratified in the U.S., perhaps fewer digital literacy or other technology programs. Or perhaps the Internet opens up the opportunity for silenced voices to be heard in the UK more than in the United States. The UK is smaller and more homogenous, which may also lead to less inequality if there are fewer silo-effects to encourage more online communication. Simply, between-country differences require more in-depth analysis than Blank's article provides.

Conclusion

'The important point is that the Internet is not always and everywhere the same and it should not be treated as if it were', writes Blank (2013), 'Theories must become more sophisticated' (p. 10). His article is in need of this very prescription with more rigorous theoretical explanations for some very interesting findings.

Ultimately, despite issues with his interpretations, Blank's research on digital production in the UK and other scholarship on the United States use similar techniques. His methodological arguments for our different findings fall short. Instead, to advance our understanding of the reproduction of inequality with online content production, we need more theoretically and empirically sound explanations to explain potential differences between countries. We also need

more rigorous studies to test his conclusions. Otherwise, our understanding of stratification with digital technology, particularly around content creation, is limited. Blank directly challenges Hargittai (2008) when he says that his findings 'provide some support for the idea that content creation does not always reinforce existing stratification'. Yet American scholarship shows the persistence of digital production inequality.

Digital democracy is at stake, whether for overtly political commentary or not. Given the increasing importance of the Internet for civic life, we need to fully understand both the consumption and production of all digital content. Digital inequality and online content could be better understood by examining online content itself, but being careful not to rely exclusively on Big Data, which exclude those not online.

Notes

1   Hereafter Hargittai and Walejko (2008) will be referred to as H/W and Correa (2010) as Correa.

2   Blank also references H/W's Table 3 in his Note 7, but this is in reference to descriptive statistics.

3   To further address the artefact question, I isolated one study that had many of the digital production activities (dependent variables) under study and was able to test a variety of digital production activities in relationship to class and also found inequality.

4   The naming of the three categories also reflects the conflation between content and practice. For instance, 'skilled content' does not describe the content itself, like 'political content' but describes an attribute of what the author guesses is required of these variables over the others. Yet oddly, posting videos, which requires quite a bit of skill, as well as can often be more 'creative' than 'social' is left out of this category. Furthermore, Blank ends up classifying and comparing Hargittai's content variables of music, poetry/fiction, film/video and artistic photography with his 'social and entertainment' category, rather than with the creative content variables. Blank does sometimes refer to his study of different types of content creation, rather than content, but he nonetheless emphasizes that these activities result in different types of content.

References

Alwin, D. & Hauser, R. (1975) 'The decomposition of effects in path analysis', *American Sociological Review*, vol. 40, no. February, pp. 27–47.

Blank, G. (2013) 'Who creates content? Stratification and content creation on the Internet', *Information, Communication & Society*, vol. 16, no. 4, pp. 590–612.

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*, vol. 16, no. 1, pp. 71–92.

Hargittai, E. (2008) 'The digital reproduction of inequality' in *Social Stratification*, ed. D. Grusky, Westview Press, Boulder, CO, pp. 936–944.

Hargittai, E. & Walejko, G. (2008) 'The participation divide: content creation and sharing in the digital age', *Information, Communication & Society*, vol. 11, no. 2, pp. 239–256.

Kerckhoff, A. C. (2001) 'Current of thought: sociology of education at the dawn of the 21st century', *Sociology of Education*, vol. 74, no. Extra issue, pp. 3–18.

Robinson, L. (2009) 'A taste for the necessary', *Information, Communication & Society*, vol. 12, no. 4, pp. 488–507.

Schradie, J. (2011) 'The digital production gap: the digital divide and Web 2.0 collide', *Poetics*, vol. 39, no. 2, pp. 145–168.

Schradie, J. (2012) 'The trend of class, race, and ethnicity in social media inequality: Who still can't afford to blog?' *Information, Communication & Society*, vol. 15, no. 4, pp. 555–571.

Jen Schradie is a doctoral candidate in the Department of Sociology at the University of California, Berkeley. She is affiliated with the Berkeley Center for New Media and the Data and Democracy Initiative. Her research focuses on the intersection of social media, social class and social movements. Address: Department of Sociology, UC Berkeley, 410 Barrows Hall, Berkeley, CA 94720–1980, USA [e-mail: schradie@berkeley.edu | website: www.schradie.com | twitter: @schradie]