

Proposal

Mukund Sundararajan

November 1, 2021

Introduction and Research Question

As we continue to live with the effects of the COVID-19 pandemic, the concept that technology would be used as a tool in education is non-negotiable. As classes have been forced online, it has become more of a necessity than ever to have high speed, reliable internet. However, it would be dismissive to view access to quality internet as a mere temporary replacement for in-person learning. It is easy to imagine that access to internet should allow students to search for information quickly to better their understanding on any subject, but perhaps it serves more of an entertainment role for most students rather than educational. While many students report that they regularly use the internet to complete their assignments or study for exams, it is not clear whether this is an adaptation to the availability of the internet or an extension on traditional learning.

I want to examine to what extent both access to internet as well as the speed of the connection have on educational outcomes for secondary students in British Columbia. If there is a significant effect, the primary goal then is to establish regions that would benefit from greater investment and identify the minimum connection speed to ensure all students are performing on an even playing field. The secondary objective would be examine the persistence of the rural/urban digital divide by looking at the difference in educational outcomes as a result of connection speed in these two communities.

Background

Studying the effects of broadband proliferation on key social indicators is not a novel idea. Since its introduction, the internet has fundamentally changed how we live, and its effects can be seen through many metrics. For example, a simple search pulls up studies inspecting the effect of the internet on economic growth (Choi & Hoon Yi, 2009), international trade (Lin, 2015), and interest rates (Luo, Li, Peng, & Fan, 2018), among many others. Unsurprisingly then, the effect of internet usage on student performance has been studied extensively.

Studies in this area primarily focus on the relationship between internet usage patterns and academic performance. In their paper, Austin and Totaro (2011) establish that high school students who are exposed to moderate internet use at home outperform their peers who either don't have such access or show intense internet usage at a statistically significant level. Along similar lines, Xu, Wang, Peng, and Wu (2019) and Siraj et al. (2015) show that frequent internet usage is positively associated with undergraduate student academic performance. This gap is particularly stark when compared with students who have no access or access only via a cell phone due to living in rural areas. Such students are likely to perform worse not only on exams but also on homework completion and are less likely to attend university (Hampton, Fernandez, Robertson, & Bauer, 2020). Thus it is clear that access and moderate use of the internet has some impact on student performance.

The importance of providing high quality connections is backed up by policy objectives from the Government of Canada. The CRTC has a major objective to ensure that all Canadians have access to high quality mobile and broadband internet (Radio-Television, Commission, et al., 2020), supported by the Universal Broadband Fund for a total of \$2.75 billion in funding. This is adding to existing provincial initiatives, leading to much improved connectivity in rural areas (Rajabiun & Middleton, 2013).

However, much of the research focuses either on specific usage activity patterns, or the importance of accessibility. I want to add to this literature by looking at how important the connection quality is to student performance, which is not covered by existing literature.

By examining the trends in student performance correlated with improved connectivity due to public investment, the goal is to identify what the minimum quality of connection is necessary from an educational standpoint.

Data

The data comes from a few different sources and is linked together programatically, and the relevant notebooks can be accessed here on github. Broadband information for BC is a subset of the National Broadband Data published by Innovation, Science, and Economic Development Canada under Canada’s Open Government Initiative (Winters & Lacharité, 2021). This dataset consists of hexagons of around 25 squared kilometers over which bandwidth data is aggregated. The variables of interest for each hexagon are the various boolean speed markers: cartesian product of Wired, Wireless, and Combined with (speeds notated in download/upload Mbps) 5/1, 10/2, 25/5/, 50/10, and LTE availability. With this dataset, we can establish the bandwidth availability for each hexagon. This is cross-referenced with Pseudo-Household Demographic Distribution Dataset, published by the same organisation, to retrieve population counts for each hexagon (Winters & Pilon, 2020). For the sake of brevity, only the summary statistics for the combined speeds is shown here: 1, the rest can be seen in the linked notebook.

| | Combined_lt5_1 | Combined_10_2 | Combined_5_1 | Combined_25_5 |
|-------|----------------|---------------|---------------|---------------|
| count | 420594.000000 | 420594.000000 | 420594.000000 | 420594.000000 |
| mean | 0.938097 | 0.889478 | 0.922434 | 0.871981 |
| std | 0.240979 | 0.313540 | 0.267489 | 0.334111 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 1.000000 | 1.000000 | 1.000000 | 1.000000 |
| 50% | 1.000000 | 1.000000 | 1.000000 | 1.000000 |
| 75% | 1.000000 | 1.000000 | 1.000000 | 1.000000 |
| max | 1.000000 | 1.000000 | 1.000000 | 1.000000 |

Table 1: Summary Statistics for National Broadband Data

Information regarding secondary school performance is scraped from a report from the

Fraser Institute, who publish an annual report cataloguing various variables of the top 250 ranked schools in BC (Emes & Cowley, 2020). Variables of particular interest to us from this dataset is the grade 12 enrolment, average exam mark, graduation rate, ESL %, Special Needs % and overall rating. The overall rating takes into account the above variables in addition to gender gaps in academic achievements to produce an overall rating, which is tracked from 2015-2019. This is merged with a dataset from Education Analytics and published by the BC Provincial Government that contains all BC schools with indicators for French programs (*BC Schools - K-12 with Francophone Indicators*, 2020). The reason this is done is because the BC dataset contains the district number, address, and latitude/longitude of every school in BC. By combining these datasets, we can assign each populated hex to the closest school within a reasonable distance on the grid. There were some differences in the dataset, which were reconciled. The details of the transformation are available on the GitHub repository. The summary statistics for the variables of interest in the year 2019 are shown here: 2

| | %_ESL | %_Special_Needs | Avg_Exam_Mark | %_Grad | Overall |
|-------|------------|-----------------|---------------|------------|------------|
| count | 235.000000 | 235.000000 | 235.000000 | 235.000000 | 235.000000 |
| mean | 3.379574 | 11.887234 | 69.148085 | 97.065532 | 6.011915 |
| std | 4.283132 | 4.780007 | 4.837914 | 2.689002 | 1.761030 |
| min | 0.000000 | 0.000000 | 54.200000 | 86.700000 | 0.000000 |
| 25% | 0.150000 | 8.900000 | 66.300000 | 95.300000 | 4.950000 |
| 50% | 1.900000 | 11.500000 | 68.600000 | 97.700000 | 6.000000 |
| 75% | 5.100000 | 14.650000 | 71.650000 | 99.100000 | 7.050000 |
| max | 21.400000 | 29.100000 | 85.400000 | 100.000000 | 10.000000 |

Table 2: Summary Statistics for School Data in 2019

There are currently two ways to determine whether a region is rural or not: one solution is to look at the list of communities that are part of Rural Coordination Centre of BC, and set every school present in one of those communities as rural, and the rest as urban. Another solution is to look at the Postal Code Conversion File provided by StatCan (Canada, 2021) which contains a Statistical Area Classification code that sets the metropolitan influence of each postal code.

Methodology

I think the first step here is to establish a relationship between the bandwidth availability and exam scores. In order to do this, I think a regression specification with exam scores as the dependent variable, and fastest available bandwidth as a coded independent variable. Adding enrolment numbers, % ESL, % Special needs, and Rural/Urban binary control variables makes sense. It is not clear yet how to add median household income as a control as the census data is aggregated at too coarse a granularity, but that would be a good addition.

However, all this does is establish a correlation, I'm not sure how to show causality yet. One thought is to use a differences-in-differences model; trying to find a few rural schools that are on the same path both in terms of educational outcomes and bandwidth availability, and see if there's any divergence if bandwidth improves for some schools but not others. This requires some further digging through the combined data.

References

- Austin, W., & Totaro, M. W. (2011). High school students' academic performance and internet usage. *Journal of economics and economic education research*, 12(1), 41.
- Bc schools - k-12 with francophone indicators. (2020). Education Analytics. Retrieved from <https://catalogue.data.gov.bc.ca/dataset/bc-schools-k-12-with-francophone-indicators>
- Canada, S. (2021). *Postal Code Conversion File, February 2021 Postal Codes, 2021*. Abacus Data Network. Retrieved from <https://hdl.handle.net/11272.1/AB2/SOQYLN> doi: 11272.1/AB2/SOQYLN
- Choi, C., & Hoon Yi, M. (2009). The effect of the internet on economic growth: Evidence from cross-country panel data. *Economics Letters*, 105(1), 39-41. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0165176509001773> doi: <https://doi.org/10.1016/j.econlet.2009.03.028>
- Emes, J., & Cowley, P. (2020). Report card on british columbia's secondary schools 2020.
- Hampton, K., Fernandez, L., Robertson, C., & Bauer, J. M. (2020). Broadband and student performance gaps.
- Lin, F. (2015). Estimating the effect of the internet on international trade. *The Journal of International Trade & Economic Development*, 24(3), 409-428. Retrieved from <https://doi.org/10.1080/09638199.2014.881906> doi: 10.1080/09638199.2014.881906
- Luo, C., Li, M., Peng, P., & Fan, S. (2018). How does internet finance influence the interest rate? evidence from chinese financial markets. *Dutch Journal of Finance and Management*, 2(1).
- Radio-Television, C., Commission, T., et al. (2020). Broadband fund: Closing the digital divide in canada. *Consumer Information*, April, 27.
- Rajabiun, R., & Middleton, C. (2013). Rural broadband development in canada's provinces: An overview of policy approaches. *Journal of Rural and Community Development*, 8(2).

- Siraj, H. H., Salam, A., Hasan, N., Jin, T. H., Roslan, R. B., & Othman, M. N. B. (2015). Internet usage and academic performance: a study in a malaysian public university. *International Medical Journal*, 22(2), 83–86.
- Winters, D., & Lacharité, Y. (2021). *National broadband data*. Government of Canada. Retrieved from <https://open.canada.ca/data/en/dataset/00a331db-121b-445d-b119-35dbbe3eedd9> doi: NBD.7.2
- Winters, D., & Pilon, P. (2020). *Pseudo-household demographic distribution*. Government of Canada. Retrieved from <https://open.canada.ca/data/en/dataset/b3a1d603-19ca-466c-ae95-b5185e56addf> doi: PHH.3.5
- Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 98, 166-173. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0747563219301554> doi: <https://doi.org/10.1016/j.chb.2019.04.015>