Always online, a blessing or a curse?

An empirical look into the effects of the Internet on the educational attainment of Norwegian youth

Magnus Flemmen Hellsaa

Master's thesis

This thesis completes the following degree:

Master's in Economics

University of Bergen, Department of Economics
June 2023

UNIVERSITETET I BERGEN

Acknowledgements

I would like to give a huge thank you to my supervisor Rita Ginja for all the help and guidance making this thesis possible. The amount of time and effort she has put into this project to ensure my success is greatly appreciated.

I would also like to thank my wife Jennifer for being amazing, and her unwavering support and help with everything throughout the process.

Finally, a big thank you to my family, Tore, Torill, Marie, Inger and Else-Marie, for all their help and support throughout the years.

Abstract

The accessibility of quick and reliable Internet has skyrocketed in the last two decades, and youth nowadays spend more time online than ever before. This paper will analyze the effect this change has had on the educational attainment of youth in Norway, using the expansion of the LTE-network. The staggered expansion creates a natural experiment, allowing for the use of difference-in-differences approach. In addition, the exogenous change the expansion caused allowed for the use of mother fixed effects approach.

By combining Norwegian registry data with Internet coverage, the estimates show a varying effect of increased Internet access based on approach. The mother fixed effects approach estimates the effects on middle school and high school grades to be negative. However, it does not find an effect on high school completion regardless of gender. This effect could also be caused by the birth order effect, rather than the difference in coverage.

Difference-in-differences showed a positive effect on middle school and high school grades, but the effect is lower for boys than girls. This approach also finds a positive effect on high school completion at the age of 19, and this looks to be stronger for boys than for girls. We also introduce a behavioral outcome, having a child before the age of 20, but the finding on this is not consistent enough to argue an effect.

Analysis was done using STATA/SE 17.0.

Contents

Tables and figures

1. Introduction

In recent years, there has been a massive change in the accessibility of the Internet, smartphones, and computers for the average person in Norway. With both the cost and the quality of the services changing, it is possible for practically everyone to be connected to the Internet at all times during their everyday lives. This also includes younger people and children and has led to children getting their first phone at a younger age (Medietilsynet, 2018). This has allowed the use of phones and computers to become a big part of their lives from an early age, and the society around them has grown to expect this of children as well.

The effects of this have naturally become a hot topic in the news (Eilertsen, 2018; Velsand, 2020). Whether it is parents worried about overuse (Lunde, 2023), discussions of research showing slower brain development (Kotsbakk and Toftaker, 2019) and doctors warning about negative effect on children's physical development (Gerhardsen and Hjelle, 2023), it is definitely a topic which engages and affects a major part of society.

However, there is not enough research to base the conclusions on, or to estimate how big the influence is. In the time period, there has been a drastic change in the mental health of youth. In the 1990s and the early 2000s there was a positive trend, youth felt better about themselves and had a decreasing amount of mental health issues. However, around 2010 there was a negative shift. The trend turned, and an increasing number of youths started having trouble with a multitude of social and mental health issues (Krokstad et al., 2022; Twenge et al., 2018). As there is no doubt the technology is here to stay, it is important to research the effect of it on the future generation.

One of the biggest changes in society at this time was the wide-spread accessibility and use of the Internet and of social media. The evolution of social media has drastically changed the lifestyle for many people, and continues to do so, for better or worse. While the opportunity to reach out or stay in contact with others is an important function, it has some clear drawbacks. The addictiveness of phones and computers has made a big part of the current generation of children reliant on always having it available.

There are nearly endless possibilities using this technology, which can be both a blessing and a curse. Especially for youth, it might be difficult to take advantage of the technology without falling victim to the drawbacks (AACAP, 2017). It can help children evolve through learning and connecting with a broader network than previous generations could. The access to practically any information, in a device small enough to fit in your pocket. It also allows youth to find new ways of learning, as traditional schooling serves some better than others.

However, the Internet can also be a great distraction, and cause disinterest in learning and growing as it does not give the same instant gratification as video games and social media are engineered to do. This loss of focus can cause youth to both struggle with learning and be more inclined to give up on overcoming these difficulties as they could easier gain gratification from the virtual world.

In addition to this, the mental health of youth can also take a toll from the use of social media. By seeing others' lives through the lens of social media, youth can get an unrealistic idea of how their life should be or should look like. This can worsen their mental health, which in turn can affect their educational attainment. Research supports this claim, showing that mental health is an important denominator for educational attainment and other long-term outcomes.

The case of worsened mental health and reduced educational attainment amongst youth can have consequences for both the individuals and society (Card, 1999). The two often go hand in hand and can amplify each other, as a student struggling with mental health might have problems keeping up at school. As both problems get worse, the person will not be able to follow their desired path and fulfill their potential (Cornaglia et al., 2015; Currie et al., 2010). Additionally, it can lead to the individual spiraling downwards, losing out on the potential positive effects of educational attainment on mental health (Chevalier and Feinstein, 2006). This is negative not only for the student but for society as a whole, as the person will both create less benefits in the future and be of higher cost to health services (Egan et al., 2015; Smith and Smith, 2010).

In this thesis I will study the effect of this technological change on the educational attainment of Norwegian youth. To do this, I will use different measures of educational attainment outcomes in high school-age students in different municipalities in Norway. I will be using the staggered 4G broadband expansion which started in 2009 as an exogenous change to the Internet access of the students in different municipalities. I will with this attempt to isolate the effect of increased Internet access on educational attainment.

This paper will be an empirical addition to the already existing literature. The topic of the Internet and use of screens on youth is highly relevant in many fields, and there are quite a few papers that has been published in the last few years. Most do however raise the question of a relationship between the two and dive into the consequences based on correlations, rather than attempt to find a causal relationship (Krokstad et al., 2022; Surén, 2018; Eines et al., 2019). My paper will therefore supply an attempt to find a causal relationship, using Norway's high quality registry data, which has not been done earlier.

2. Background

In this section, we will go through some of the background for the analysis. First taking a look at relevant earlier research done on the effects of Internet and social media on mental health and individual decision-making. After, we will dive into how, where and why the expansion took place, in order to get a deeper understanding of the structural background of the situation where the changes took place.

2.1 Existing literature

As social media and the use of screens has become a major part of people's lives and of society as a whole, it has also become an area of interest for research around the effects. In recent years, there has been a lot of new research on the effects on mental health and other long-term outcomes amongst youth, using data from the last two decades. However, a lot are written in the medical- or psychology fields and focus on the correlation between social media and mental health amongst youth. The results of these papers are also quite varied, and many find little effect (Valkenburg et al., 2022).

As for economic literature, there have been some recent papers focusing on finding a causal link by using an exogen shock, or an experimental approach. Braghieri, Levy and Makarin (2022) uses the launch of Facebook across American colleges and universities to look at its effects on the students' mental health. When Facebook was introduced, it initially put its exclusivity highly and only a few universities were granted access. In the following months, it was gradually made available at more and more places, allowing the researchers to use it as an exogen shock on student mental health. While Facebook is only one form of social media, it was a frontrunner in what social media is today and its effects are likely transferable to other types used today.

The paper finds a negative effect on the students' mental health, which stems from the site fostering increased negative social comparisons amongst students. Furthermore, this negative effect on mental health brought a higher likelihood of students struggling with achieving their preferred academic results because of poor mental health (Braghieri et al., 2022).

Another paper looking at the effect of social media used an experimental approach. Allcott et al. (2020) used Facebook and the US midterm election in 2018 to study the connection between social media and personal welfare. The experiment consisted of offering a small one-time payment to randomized individuals, who had to disable their account for four weeks leading up to the election. Their results are based on the people's own feelings of well-being, as well as change in behavior.

The results of the experiment are somewhat mixed. The individuals who disabled their account report an improved feeling of self-being, and that their demand for the service has been substantially reduced. This might suggest addiction and projection bias has led to an overvaluation of the product. In addition to this, the participants who disabled their account are less informed about current events than previously and read less news but has also made them less polarized. As this is a current trend in American and European politics, the authors speculate social media might play a role.

There are however some clear positive effects as well. It allows for organization and socialization, which might not be possible without the platform. It also keeps the users informed and updated on different topics, and the users put a high value on the service in general. Therefore, the authors state that the negative effects are smaller than one might expect, but big enough to be given real concern. (Allcott et al., 2020).

The method of using an Internet expansion (the broadband network – 3G) as exogenous variation has also been tested before. Bhuller et al. (2013) uses the expansion in Norway in the early 2000s to look at the effects of Internet access on sex crime. They find the expansion increased Internet use, and that this change had a direct, positive effect on sex crime propensity. They also discuss the avenues the Internet can facilitate for such crimes, for example through instant messaging and social networking, which closely resembles today's social media. I will come back to this later.

Another paper which used the broadband expansion in Norway is Hvide et al. (2022). Similarly to Bhuller et al. they used the expansion in the early 2000s, using it to analyze the effects of individual investors' behavior on the stock market. They used the expansion as it allowed them to use a change which affected a broad group of people, not just those who already invest. Their results were that the expansion led to a strong increase in stock market participation, and that individuals' investment decisions were more in line with portfolio theory.

As previously mentioned, there has also been some work on the topic using Norwegian data without the attempt of finding a causal relationship. Using data from HUNT, Krokstad et al. (2022) find a shift in the mental health and general well-being amongst youth around 2010. This coincides with the introduction of, and a surge of popularity for, social media such as

Instagram and Snapchat. They argue that this is likely at least partly the cause of the change in youth's well-being and advice for reduced use and availability. The trends they looked at also indicated that this was mainly a problem for youth, as they did not find similar trends for adults. Rather, it seemed to have a positive effect on the depressive symptoms of adults aged 60 and above.

2.2 Institutional background

Before getting into the analysis, we need an understanding of the institutional background of the cellular network in Norway and how it changed over the period. As I will be using the expansion of the LTE-network, which provides a 4G cellular network, I will first explain how it was before the expansion. Thereafter, I will go through where the expansion took place and how this was decided. Now, in 2022, approximately 100 per cent of Norwegian households have basic 4G-coverage outside their door. While this coverage does not necessarily guarantee a minimum Internet speed, around 98 per cent of households have down- and upload speed of at least 30 Mbit per second. Therefore, even the basic coverage provides higher speeds and reliability than 3G did (NKOM, 2022).

The first generation of commercially available cellular network, now known as 1G, was launched in Japan in 1979, and soon followed by the Nordic Mobile Technology in Sweden and Norway. It was an analog network, connecting bigger systems such as cars, as phones were not yet handheld. It was followed by 2G in the 1990s, as the technological advancement of the now mobile phone required better cellular services. However, the demand for more than just communication such as calling and texting, such as access to the Internet, was growing.

The third generation, 3G, was the first generation which made it possible to browse the Internet. To match the growing number of phones and people's increased use of them, 3G used new technology to increase speed and capacity. While enabling browsing the Internet, it typically did not provide the speed or capacity to use all online services. Basic functions like messaging through social media was now possible using a mobile phone, while down- and uploading photos could be possible, but both slow and rather expensive. Streaming, online gaming, and other services of the like was still not possible using the cellular network. The expansion of this network in Norway was used by Manudeep Bhuller et al. in their paper on Internet and Sex Crime (Bhuller et al., 2013).

While this is similar to how the cellular network is today, the demand for speed and capacity outgrew what the technology could provide. As it was used in both phones and other portable devices like laptops, it was not enough to just enable access to the Internet. The network had to have the speed and capacity to handle down- and uploading, streaming and more. Therefore, in late 2009, the expansion of the LTE-network began. This network had the capability to provide 4G, the next generation of cellular data. With access to 4G, handheld devices can browse any type of social media, streaming sites, online gaming and much more, services which earlier were very slow or not possible at all with 3G.

I will be using the expansion of the LTE-network as a measure of Internet access. At the time the expansion started practically all people in Norway, regardless of municipality, had basic broadband coverage. This basic coverage required a speed of at least 640 kbit per second download speed and 128 kbit upload speed, using a home network. More than 95% were also covered by the cellular network, giving access to the same Internet speed outside as the broadband at home. In addition to this, houses in central municipalities are offered higher speeds for home broadband. Around 50% of households had at least one provider offering speeds of 25 Mbit per second. The cellular network could still only provide 3G-coverage, which only guaranteed the basic speed previously mentioned. An important distinction is that both home and cellular broadband can be based on the cellular network, meaning the expansion would not only increase the access to Internet services outside or on a phone but also the normal broadband in the household.

The expansion would therefore offer increased down- and upload speed. Before the expansion the basic coverage guaranteed download speed of 640 kbit per second, and upload speed of 128 kbit per second. After the expansion, the basic coverage guarantees 30 Mbit per second of both, meaning the download speed became almost 50 times faster and the upload speed over 200 times faster. This allows the user to access a lot of new media, which was not previously possible. In the world of social media, using 3G is sufficient for text-based services such as Facebook. However, most social media popular amongst youth is based on pictures and videos. As such, the upgrade to 4G allows the use of these types of social media not previously possible.

The next step is to look at how the expansion took place. To distribute the cellular network, there needs to be some physical infrastructure in place, as well as radio frequencies to distribute the signal. As the new 4G-network was to upgrade the former 3G-network, the necessary infrastructure could be achieved through upgrading the old. However, the radio frequencies which were needed to distribute the signal were a bigger challenge as it was to a large extent taken up by the traditional analog distribution of television signals.

At this point in time, most televisions were no longer using the analog signal as it was replaced by digital-, cable- or satellite-TV. Therefore, it was decided that the analog distribution was to be shut down, and this was done between March of 2008 and December of 2009. While this opened up the possibility of high-speed cellular network, it could also allow for a drastic increase in the digital television signal. As the Norwegian government has ownership over the radio frequencies, it sells the rights to use them through auctions. Through auctions, the Norwegian government allow private companies to use the network, while still being the owner. This generates income for the Norwegian state without selling and allow private companies to broaden their reach.

The Norwegian Ministry of Transport decided in September of 2009 that the radio frequencies were to be used for cellular network, rather than be used for new, higher quality television signals (Bjørkeng, 2009). However, before the auction could take place there was a long process to go through, setting all the rules and requirements for how the auction and use of the frequencies would be. This process took over four years, and therefore the auction did not take place until December of 2013. While the use of these frequencies was integral to the distribution of the network on a national level, as they had very long range, they were not necessary for distribution of 4G in smaller areas. Therefore, the companies providing the service were unable to expand nationwide until after the auction. However, they could still use the radio frequencies they already used for 3G to provide 4G for smaller areas, as the frequencies had the capacity, just not the range.

In the meantime, the companies providing the cellular network could use the radio frequencies already in use for 3G, by upgrading to the LTE-network. Late in 2009, Oslo became the first area in Norway to receive 4G-connection, using these frequencies. At the time, NetCom was the only company which was able to provide the service in Norway. As their main competitor Telenor were not going to offer 4G-connection until 2012, their LTE-network was quickly expanded to cover more major cities. In the next year, NetCom expanded their upgraded network to Bergen, Stavanger, and Trondheim. By early 2011, people in Bergen, Trondheim and Stavanger could start connecting to 4G, still using the radio frequencies used by older cellular networks. After the biggest cities had coverage, the expansion continued to the areas around the cities, as well as popular travel spots where many cabin-owners were eager to access better Internet-connection, such as Hemsedal and Trysil.

However, this expansion did not include the use of 4G on mobile phones. It could only be used on bigger devices such as laptops, using an extra modem connected to the device. The 4G-network first became available for mobile phones in late 2012 with Telenor, and in mid-2013 with NetCom. This allowed even small devices to seamlessly connect to the network, giving all smart-phone users access anywhere the expansion had hit. In addition to this, the devices themselves were capable of switching to other cellular networks such as 3G if it lost connection to the 4G-network, giving a very consistent and reliable connection.

The problem with using the radio frequencies they had been using for the old cellular network was the reach of the signals. As the frequencies were relatively high, they are good at sending high amounts of information over short distances. However, the signals lose strength quickly when traveling, and would therefore need more physical infrastructure such as radio towers than a low frequency. Expanding to the districts would therefore be very costly and inefficient, compared to using a lower frequency to provide the signal. For this reason, the LTE-network was hardly expanded outside the cities and high-density cabin areas before the auction, as they could reach far more customers for each radio tower. The frequencies being auctioned out were high enough to be used for cellular data service, while low enough to be able to travel much further. As the preparation for the auctions took so long, the companies were ready to expand as soon as the frequencies were made available for them. This ensured a very rapid expansion after the auction, as the companies had already upgraded most of their network all over the country while waiting for the Ministry of Transport to complete the auction.

The auction was done through secret bidding, and the results were announced on the 6th of December. There were three companies which were able to buy around a third each, Telenor, NetCom (today known as Telia) and TelcoData (today known as ICE). One of the blocks of frequencies, which was considered to be the best for the cellular network expansion, came with a requirement of good cellular network coverage for 98% of the Norwegian people within five years. NetCom won this block of frequencies, which ensured their expansion would cover all of Norway as soon as possible. As their main competitor, Telenor had to put a lot of resources into matching the nationwide expansion as well, so as to not lose their market share. This would ensure both coverage and competitive pricing for most Norwegians within a few years.

I will be using the upgrade which allowed mobile phones to connect to the 4G network as my indicator of Internet access. This is a better indicator than the release of 4G as it can be used on mobile phones and does not require another modem to be used. For most areas this is not relevant, as the municipalities which got access to the LTE-network after mid-2013 got the mobile phone-compatible version of 4G from the start.

3. Expected effects of expansion on educational attainment

The availability of the 4G-network on mobile phones is what made it readily available for youth. It no longer required a computer and home or school WiFi to connect to social media, streaming services and much more. Youth could now connect to all of this through their mobile phones, regardless of where they were or what they were doing. This is likely attributed to the increased use of Internet amongst youth, which has been happening for the past decade.

The increased speed and reliability stemming from the expansion has allowed for more efficient use, which could therefore lower the amount of time spent. However, this does not seem to be the case, as youth spend vastly more time online now than what they did a decade ago. The bettering of the service has increased the value of what the user gets out of it, as less time is spent waiting and more time is spent doing what the user finds interesting. This also lessens the barrier to entry, so to speak, as loading in takes but a few seconds, meaning it can be used in the smallest of downtimes. The reliability of it also allows constant use without any problems, which can make youth feel obligated to be always on and available. The increased use has also been matched by new creations and things to do using the service. A person could spend their entire life browsing around the Internet, and never run out of sites to discover and content to consume.

However, the expected effect of increased use is not limited to the effects of increased use of social media or video games. This technological evolution has also brought new tools and ways to solve problems, as well as making working together online easier and better. Therefore, the increased use of screens amongst youth is not only based on non-schooling activities. This expansion will likely increase the use of screens in school, for schoolwork and in youth's pursuit of other interests as well as social media and video games. Therefore, the LTE-network expansion cannot be simply looked at as an increased use of one thing, whether that is schoolwork, social media, or gaming. There will be many different mechanisms at work, and the expansion will affect people differently.

The increased use of many online services can be partly blamed on addiction. While it is not all-consuming or inherently something that must affect your whole life, the use of these services is typically much higher than they need to be. An addiction will lead to spending too much time using the service, but also inflate the sense of worth the individual has of the product (Allcott et al., 2020). In addition to this, it can take up more time than just the time spent on the service. As previously mentioned, social media can seem to have a negative impact on mental health, but it can also make youth think about it even when not using it. This could lead to youth being distracted, even in circumstances where they are not able to use the services such as during class.

There has also been some empirical research on the problem of digital addiction. Allcott et al. (2022) attempted to use a field experiment to quantify the effects of addiction to social media and screen time. Through surveying 2000 Americans and recording their smart phone screen time, they found evidence suggesting social medias are habit forming, and that people are partially unaware of their own building addiction. Using a model, they estimate self-control problems to cause 31 percent of social media use.

Onto the different services youth typically use. First, the often-hot topic and part of the motivation for this analysis, the effects of social media. Social media has in the last two decades evolved from a few basic services mostly based on text, to now covering any and every type of media used. This change has made it more necessary to have faster and more stable network connections to use social media. While loading a text didn't take long regardless of 3G or 4G, accessing pictures, videos, streaming and so on was not possible before the upgrade to the network. Looking at the types of social media Norwegian youth uses today; most are based on videos and pictures and the accessibility is affected by the quality of the network used. The use of Instagram and Snapchat relies on good Internet to allow the user to see other's content and take pictures and videos themselves. Likewise, Youtube and Twitch demand a fair amount of data in order to watch videos and livestreams. For all of these services, speed and reliability are absolute key. (Medietilsynet, 2008, 2020; Knapstad et al., 2018; Sivertsen et al., 2022)

Considering this, the accessibility of a better network would likely increase the use of social media. This increased use allows a more active online social life, both by keeping in touch with people they already know and getting to know new people. By having an easier time keeping in touch with friends, especially those who they cannot physically meet up with, it could have a positive impact on their mental health. Getting to know new people can broaden their horizon and motivate learning beyond traditional schooling. Social media has made it easier than ever to make friends online as well, making this a very relevant topic.

The increased use of social media can bring some negative effects as well. Social media both opens for and encourages constantly comparing yourself and what you have to others. Whether it is about how you look, what you do, or how you live, everything can be compared to impossible standards created by many people's aims of appearing perfect online. The effect of this drive to achieve the unachievable can be negative on both youth's physical and mental health, as well as their time management. An example of this could be youths, comparing their looks to celebrities and striving to reach a physique not possible for most people. This can be a contributing factor in youths' mental and physical health.

Social media becoming the outlet for a lot of social interactions can also bring a heavier reliance on social image and how one appears on the surface. The focus on social image can cause individuals to stop acting in their own interest, and rather act in order to portray themselves better online. An example can be made using the findings of Bursztyn and Jensen (2017). They find that while ability to do well is rewarded, effort is stigmatized, leading to individuals making less effort than what is in their best self-interest. Furthermore, this can lead to a worse self-image as the individual is not able to achieve what they wish to, as a result of the lower effort. For youth, this problem can be quite apparent. While doing well in school is typically viewed positively, putting in an effort might be viewed negatively, leading youth to lower their effort for the sake of their social image.

In addition to social media, playing video games online is another time-consuming activity which affects youth. Similarly to social media, online gaming have become more mainstream and is now a common hobby. As both are quite time-consuming activities in front of either a phone or a computer, youth are likely to spend more time on one or the other. The effect on youth mainly playing online games would likely be different from those who spend their time on social media. The increased availability and quality would be more noticeable, as online gaming is more reliant on good and stable Internet-connection. As many video games are reliant on their player base staying interested in the game, they are made to keep players playing, and reward time spent. Granting easily achievable gratification, the games have an easier time both obtaining and retaining the interest of youth, as schoolwork and the like typically require higher effort for less feeling of satisfaction. This will likely lead to both an increase in time spent gaming and take away time from schoolwork, directly impacting youth's educational growth.

While the use of online gaming as a choice of leisure is no problem in itself, it can quickly end up becoming too time-consuming, and serve as a distraction from other things. Taking the aforementioned prevalence of digital addiction into consideration as well, this can lead to many youths spending high amounts of time on video games rather than spending it more productively. It can also affect the youth's ability to focus, especially with the introduction of high-speed Internet, smart-phones, and laptops in the classroom. This compulsive, difficult to manage, use of the Internet also seems to have a stronger association with online gaming than other similar uses of the Internet such as chatting and social networking (van Rooij et al., 2010).

This can also be a lead to a potential gender gap in the effects of the expansion. While there are young women who spend their time engaged in online gaming, the vast majority of gamers are young men. A paper studying men in the ages 21 to 30 and their relationship to online gaming finds it has a significant effect on multiple aspects of their lives. Online gaming has become the main choice of leisure amongst young men, and they are choosing this over other recreational activities. In addition, it has started to impact the labor supply of these men as they would rather spend their time playing video games rather than pursuing work-related activities (Aguiar et al., 2021).

The negative effect might swing the other way as well. Using a broadband Internet expansion in Germany, Golin (2022) have found effects suggesting young women's mental health suffer more than that of men. She further suggests the effect is amplified by the intensity of Internet usage considering the effects are concentrated on young adults. Furthermore, Fletcher (2007) finds young women are more likely to have their depressive symptoms affect their educational attainment, especially on their high school completion rate and college enrollment. These findings suggest we could find gender-differences in the effects on educational attainment, and that the differences could be different when using schooling grades and school completion as the educational outcome.

Naturally, there are also youth who do not spend a significant amount of time on either social media or online gaming. How they will be affected and through which channels can again be different from those previously mentioned. If they do choose not to spend much time using the Internet already, it will increase the quality of the services they do use without bringing much negative effects. However, if they have previously been constrained from using online services, the improvement of the Internet could lead to them spending more time online.

The state of the Internet before the expansion could also be relevant in this case. While more central parts of the country had high-speed, reliable Internet in households and schools, more rural areas might not have had as good connections. This results in the expansion mostly improving the quality and availability of the Internet away from computers and the home, and not necessarily bringing new opportunities in use. As such, services such as online gaming were already available, and the expansion only allowed it to be always accessed rather than just at home. In more rural areas it could also improve home- and school network, allowing for new avenues of use than previously possible. If this is the case, it could cause those who previously did not spend much time online to be affected similarly to the earlier discussed groups of youth.

There are also some definite positive effects of the technological advancement. The accessibility of the Internet made it much easier to use for solving problems and giving youth more avenues to find help outside of school. In addition, it will help youth see problems from a different angle and open their minds to different ways of problem solving. This can be especially helpful for students who struggle with the standardized and relatively rigid ways of solving problems taught in school. It will also allow students the freedom to learn about topics not covered in school, and potentially figure out what really interests them.

Researching the effects of media censorship, Chen and Yang (2019) find in their field experiment that free use of Internet to bring multiple positive changes. First, it seems to bring a broad, substantial, and persistent positive effect on knowledge. In addition, it encourages increased acquisition and social transmission of information. As the introduction of 4G-network allows youth to access the Internet more freely, away from home and school, the effects on them could be similar and motivate them to learn more outside of school.

Looking at the introduction of new media, similar findings have been made. Research using the introduction of newspapers shows that giving youth easier access to more information and news increased their interest in politics and increased election turn-out (Gentzkow et al., 2011). The increased Internet access can affect youth similarly to the introduction of a newspaper did, as it will allow them to discover new and different information. As the availability of newspapers is low and typically based on parents, the freer and more available Internet can allow them to paint a broader picture and to learn of things from outside their area.

The availability of resources outside of school does require a higher level of scrutiny as to the quality of the source. This can be a problem, as youth might not be able to identify whether a source is good or bad, allowing wrongful information to be spread. However, if youth are taught to be skeptical about what they read online, it helps them learn the valuable skill of source criticism. This skill will not only help them in the search for knowledge online, but also in future schooling and work situations.

In their research on millennial-style learning, Carlin et al. (2016) finds that online video content affects individual’s decision-making, and the quality of the content can affect the willingness to share with others. They also find distractions such as advertisements effects how much time and effort individuals spend finding good sources, and this can lead to individuals finding lower quality content and make worse decisions. In addition, giving actionable advice with good instructions helps individuals make better decisions. However, it may decrease sharing unless it is perceived to be sufficiently useful.

The skill of being able to separate a good source from a bad is therefore highly important to teach youth, as the consequences of trusting bad sources could be rather large. Research on the topic shows the average person is both bad at detecting lies and think they are good at detecting lies. This combination can naturally lead to misinformation being spread without most realizing it. In addition to this, people put more trust in sources of information already shared by someone else. However, this in reality does not make a source any better as there is no guarantee the person who originally shared the information would accurately decide whether or not it was credible (Serra-Garcia and Gneezy, 2021).

There also seems to be a positive effect when using computer equipment in school. Using access to a new technology-aided instruction program, where access was won through a lottery, Muralidharan et al. (2017) finds clear positive effects on the students given access to the program in as little as four and a half months. This effect was also stronger for academically weaker students, as it allowed them to learn at a closer pace to their stronger achieving piers. They suggest well-designed use of technology in instruction programs is cost effective, both monetarily and timewise.

Similar findings were made by Bianchi et al. (2022) when looking at introduction of technology in an outreach program in China. By connecting students in rural areas to great teachers through the use of technology, they find the students do better on both short- and long-term basis. They find that the effect endures, by observing effects up to ten years later and argue the use of big-scale technological interventions can help reduce the educational gap between rural and urban areas.

There are many mechanisms at work when looking at the effect of Internet access on educational attainment, some more directly impactful than others. As previously discussed, the improved Internet access can affect youth directly by offering more alternatives of what to spend time on, or it can open previously unavailable avenues of learning allowing different people to learn in ways which better fit them and their skill set. However, there are many important distinctions to be made if we are to estimate the effect on youth in general.

There might also be very direct consequences. Research on American college- and university level education find a strong negative effect of classes being taught online. Figlio et al. (2013)

ran an experiment where they randomized whether the students were to attend class in-person or do the same class solely online. Their experiment showed a clear negative effect on the students who only attended class online compared to those who attended in-person. They warn that rushing to fully online classes for the sake of saving money might come at more of a cost than one expects.

Alpert et al. (2016) builds on this using a similar experiment; they add a third option which is a mix of online and in-person. While they find similar negative effects to that of Figlio et al. (2013) when comparing fully online and fully in-person, the results of the option to mix are interesting. They do not find that the students who have a mix do statistically worse than those who attend fully in-person. This suggests a potentially promising opportunity, allowing for economization of the teaching resources without sacrificing the learning outcomes of the students.

In addition to variation which does affect educational attainment, it is important to remember other changes happening at the same time. If there is an educational reform in an area around the time they receive the LTE-upgrade, we must make sure not to conclude that all change is credited to one or the other. While separating these effects is difficult, we cannot trust our estimates are accurate unless we make such a distinction. To enable us to do so, we need to use empirical approaches which will let us isolate the effect we wish to estimate.

4. Data

The data I will be using for the analysis is a combination of registry data and data on the expansion of the LTE-network in Norway. The sample will consist of the children of a random 25% of mothers who gave birth between 1988 and 2010, due to one of the empirical approaches I will be utilizing in this analysis. I will only be able to access 25% due to data access restrictions. In total, the sample will consist of 506062 individuals from all over Norway.

The registry data is compiled from several Norwegian administrative records. The demographic and socioeconomic information is from registers covering information the Norwegian population, which includes both fixed information such as birth year, gender, whether an individual is born in Norway, for both the individuals and their parents, and time-based information such as municipality of residence at the age of 13. The data on educational outcomes is from the national registry on education, including information on middle school grades, high school grades, high school completion, and participation in higher education. In addition, it has information on parental education in the form of their highest level of education achieved. The data on birth variable is from the Medical Birth Registry of Norway, including data on the individual at birth, on the mother around the pregnancy and the circumstances around the birth. All the information is connected by a unique identifier, and as the data is from national registries, there is minimal chance of measurement error caused by misreporting.

This data has been combined with yearly data on the expansion of the LTE-network in Norway, given on municipality-level. The data on the coverage in Norwegian municipalities was provided through yearly reports written for the National Communications Authority. This data will be used to estimate the Internet-use of the youth growing up in the different municipalities based on their access to high-speed Internet in the form of 4G.

4.1 Sample restrictions

For the analysis my sample will consist of individuals born between 1988 and 2010. I will base their access to 4G on their municipality of residence at the age of 13, as their access at this age is a good baseline for their access throughout middle school and high school. This group will therefore consist of some who grew up before the expansion, some during the expansion, and some after the expansion. While there are 506062 individuals in the full sample, 194974 has to be dropped due to missing data on municipality of residence at the age of 13. This can be caused by the individual not having turned 13 at the time the data was gathered in 2020, or that the individual was living abroad at the age of 13. This leaves us with our final sample which will consist of 311088 individuals.

4.2 Descriptive statistics

Before diving into the analysis, we will look at some descriptive statistics to better understand our sample and our variables. For our outcome, educational attainment, we have access to information on grading, participation, and completion of schooling at different stages of the individuals' lives. We will be using three different measurements: the individual's average grade when finishing middle school, average grade when finishing high school, and whether they completed high school at the age of 19. While the variables for grades are named GPA, the data is standard Norwegian school grading on a scale from one to six, where one is fail and six is the top score. We will be using high school completion at the age of 19 as all high school students following the standard course of study will be finished at this age. While participation and completion of higher education would be interesting to include, the LTE-expansion is too recent, so the youth affected by it are not yet old enough for this to be included.

To measure the network expansion, we will be using data on coverage in all Norwegian municipalities in the years around the expansion. This data is a percentage, based on the share of households in the municipality that has at least one Internet provider which offers access to the LTE-network. If we had individual data on access to 4G, it would be equal to either zero or one, if the individual had access or not. However, this data will provide the next best thing, which is the coverage rate of the municipality the individual lived in. As most municipalities will have a coverage rate between zero and one during the expansion, we cannot know exactly when individuals obtained access. Therefore, we will be using the coverage rate of the municipality as the value of the variable for the individual. In addition to this, we will be including a dummy variant which will be equal to one if the municipality had been hit by the expansion and therefore have a coverage rate of more than zero, regardless of the exact coverage rate, and zero if otherwise. Lastly, as we are looking at the affects the expansion has had on educational attainment, we will be using the municipality the individual lived in at the age of 13 as this is at the beginning of the formative years, where the individual likely will be using technology and will be affected by access to the LTE-network.

In addition to the main variables in our analysis, we will be including variables on individual characteristics and socioeconomic status. The variables on individual characteristics will include gender, whether the individual was born in Norway or not, and number of siblings. Variables for gender and whether an individual is born in Norway or not will be dummy variables, as these are typically perceived as either of two options.

The variables on socioeconomic status will cover both parental and municipality characteristics. Variables on parental characteristics will include parental income, education and whether they were born in Norway or not, all individually given for both parents. The income is given by which quartile of the Norwegian population the parent is in. From this, we will be using a dummy variable for high income families, equal to one if both parents are in the upper 50% of income and zero if either or both are in the lower 50%. Parental education is given on a standardized level basis, based on their highest completed level of education. The scale is from 0 to 8, where 0 is no education or finished preschool and 8 is doctorate or researcher. I will also create a dummy variable based on this data, parental college degree, which will be equal to one if either of the parents has at least level 6, requiring a bachelor's degree, and zero if otherwise. Lastly, whether the parents are born in Norway is also a dummy variable, similar to that of the individual. We will also control for municipality characteristics throughout the analysis.

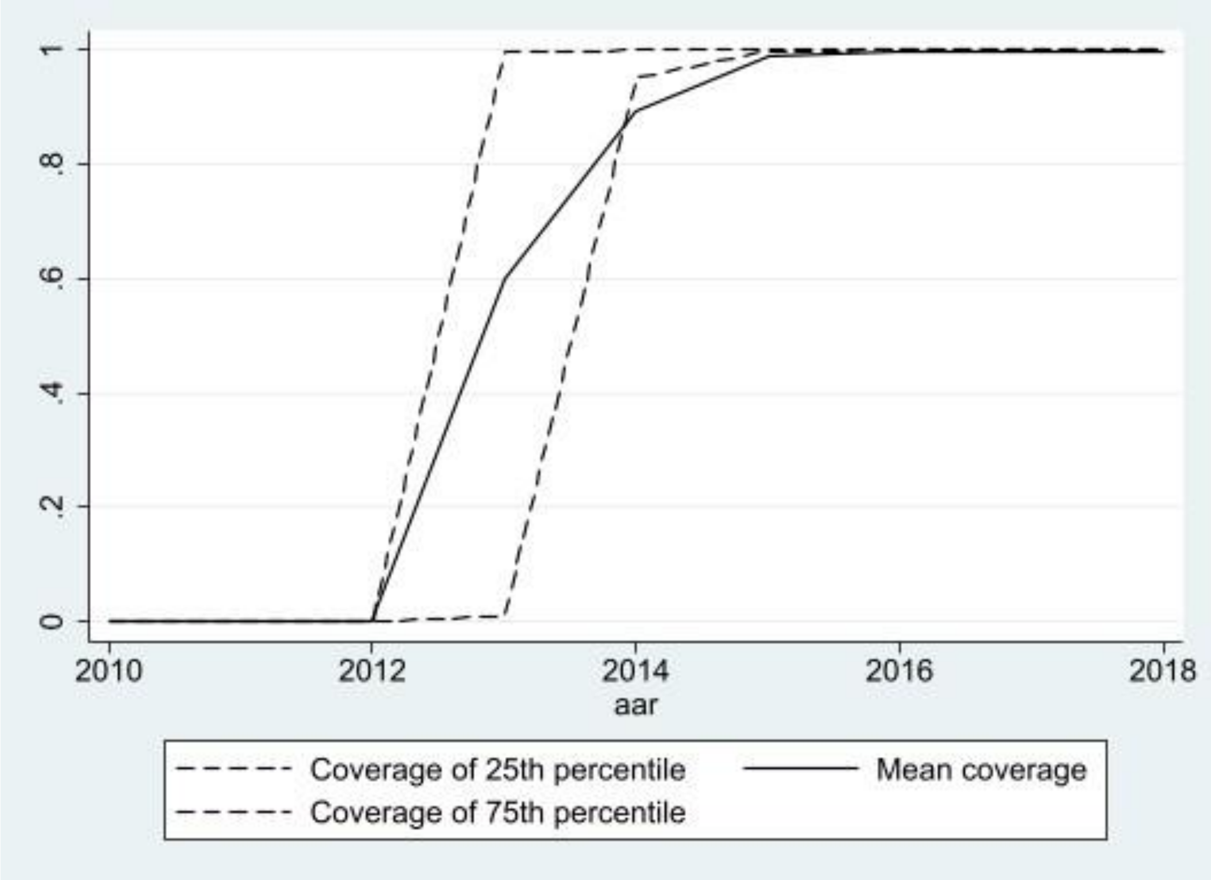
Finally, we can have a look at an overview of our sample. In the table below, we can see some of the main variables covering individual characteristics, coverage, and parental education, with number of observations, mean and standard deviation. Important to note is the fact that some individuals do not have a value for some variables, as we can see based on the number of observations. While all individuals have a registered gender and were born in Norway or not, all did not complete middle school and high school, resulting in missing variables on the educational attainment variables. It is also possible some individuals in the sample are too young to have completed high school.

Table 1: Descriptive statistics

Variable	Observations	Mean	Standard deviation
Male	311088	0,51	0,50
Share born in Norway	311088	0,79	0,41
Middle school GPA	224520	4,01	0,83
High school GPA	221511	3,96	0,87
Completed high school at age 19	201017	0,51	0,50
Coverage	278957	0,30	0,45
Coverage-dummy	311088	0,39	0,49
Highest completed education, mother	307472	4.54	1.72
Highest completed education, father	300279	4.38	1.68

Lastly, we take a look at how the coverage changed from the expansion, in the form of a graph. We discussed earlier how the expansion took place gradually, and from the graph below we can get a visual picture of how the change in LTE-network coverage was in Norway. It is important to remember that this is the LTE-network coverage, providing 4G-Internet. This graph does not reflect on other forms of Internet access, or the speed of the Internet available other than fulfilling the minimum requirement of an LTE-network. In addition, the coverage is measured and reported once a year, giving us one data point per year and no change during a year. This results in large changes from one year to the next.

Figure 1: Coverage in Norwegian municipalities



On the horizontal axis, we see the year the coverage is measured, while the vertical axis shows us the coverage rates on municipality-level. While we have more years included in our sample, the graph only shows some years for the sake of viewability. In order to give a broader view of the situation, there are three graphs included showing the 25th and 75th percentile in addition to the average coverage. These show a similar picture and reflect what was previously mentioned in this analysis. As soon as the expansion was allowed to take place, the coverage rates shot up in only a few years. Municipalities in central areas had coverage rates quickly going from zero to one hundred per cent, while it took around one more year for the expansion to reach more rural areas. Within three to four years all municipalities in the country had close to complete coverage of basic 4G. Interestingly, we also see the 75th percentile was higher than the average in 2014. This is caused by the massive difference in Northern Norway compared to the rest of the country. While the expansion had hit the most populous areas, many counties in the North had yet to get any coverage. This left a small part of the population with little to no coverage, while the majority had almost complete coverage.

Connecting this coverage to the individuals in our sample, we must remember that we base the Internet access of an individual on their municipality of residence at the age of 13. The graphs therefore show us that all individuals born before 1999 had zero coverage, and their coverage variable will be equal to zero, while all individuals born after 2003 will have complete coverage and a coverage variable close to, or equal to, one.

5. Empirical strategy

An experiment of randomizing who gets Internet access, and who does not, is not realistically possible to complete. The empirical strategy will however be to imitate such an experiment, using the gradual expansion of reliable, high-speed Internet in Norway. By using the differences between Internet access in municipalities, we can compare the changes in educational attainment in different areas in an attempt to isolate the effect of the expansion.

5.1 Ordinary Least Squares

In the initial stages of the analysis, we want to make some estimates which we later can use as a reference point. This is important in making sure we have a general idea of which direction our estimates will be, to not draw conclusions solely based on one analysis. To do this we can use an Ordinary Least Squares (OLS) regression. By running some OLS regressions using our numbers on coverage and our educational attainment variables, we get estimates on the effect of change in coverage on educational attainment. We can also include different control variables in the regressions, which will allow us to control by unobserved determinants of educational attainment.

The inclusion of control variables makes the estimates less biased as they will let us control for other differences in individuals which can impact the outcome. Therefore, it will compare the individuals on more equal grounds than if they were not included. While still not enough to claim a causal relationship, it will give us a good look at the correlation between the variables and let us have something to compare our future results to.

We can consider an example; our OLS regression shows us that treatment variable, coverage, has a positive correlation with our outcome variables, educational attainment. While we cannot these estimates to conclude that an increase in 4G-coverage causes a positive change in educational attainment amongst youth based solely on this, we can use it as a reference point for our next approach. When we then attempt to isolate the effect through a different empirical approach, we can both consider whether it is positive or negative, and how it compares to the OLS-estimate. In addition, we can use OLS regressions to look at other relevant trends, such as change in educational attainment year by year, to consider other causes of change. If educational attainment is increasing every year, the positive correlation can be caused by both variables positively correlating with time.

To run these regressions and estimate the effect of Internet access on educational attainment, I will be using a linear approach. Below is the initial equation, to which different control variables will be gradually added. Doing this, we can both get a better estimate of the effect we are looking at, and a better understanding of other factors affecting the estimates.

$$Y_i = \alpha + \beta Cov_i + \varepsilon_i \quad (1)$$

In the equation, Y_i represents our dependent variable, or outcome variable, which will be either the educational outcomes of high school completion at the age of 19, GPA in middle school, or GPA in high school, or having a child before the age of 20. These four are all important indicators of educational attainment, and we will run the analysis using each of them as the outcome variable. Doing this, we make sure to get a broader picture of the effect on educational attainment, rather than a specific measure. α is a constant, allowing for the base value to be different from zero. Cov_i represents the independent variable, which in our case is Internet access. This variable will take the value of the Internet-coverage in the municipality of the individual at the age of 13. The β will therefore be our estimate of the effect, and while Y and Cov will take different values depending on the individual, denoted by the small i, β will only take one value and as it will be an estimate for our entire sample.

In addition to these, which estimate the effect we are interested in, we need to allow for Y to take a base value, and differences between individuals other than what is encapsulated in Cov and Y. The former is done through the α, which takes a common value for all individuals similarly to our estimate of β. The latter is done through the final part of the equation the error term ε_i. It contains all the variation in Y which is not explained by a change in Cov. If this was equal to zero, it would mean we can explain all the variation in educational attainment between individuals with the difference in Internet access. This is typically not the case, as there are near endless different factors affecting the outcome, but this does not stop us from getting an unbiased estimate. For an unbiased estimate, we need the error term to have an expected value of zero. This would entail our estimate being correct for the entire sample, even though there are differences between the individuals in the sample.

A common reason for the error term to be different zero, and therefore making our estimate biased, is the omitted variable bias. This bias is caused by correlation between the error term and the independent variable, in our case the Internet access, breaking the assumption of correlation between independent variables and error term equal to zero. If this is true, it will lead to a bias in the direction of the correlation. Therefore, we will not only have a biased estimate, but we do also not know whether it is biased in a positive or negative direction. This will render our estimate near useless, as we cannot argue for its validity.

A method we can use to avoid the omitted variable bias is to introduce control variables for individual characteristics, and municipality fixed effects. For the expected value of the error term to be zero, we need to control for individual variation which affects the estimates. To do this, we should have all individual variation included in our regression. While we cannot add unobservable variation to the regression, as we cannot acquire data for them, we are able to introduce control variables of observable individual variation. This is done by adding the variable we wish to control for into the equation, accompanied by a variable representing the estimated effect of that variable. The control variables will help us include individual variation in the regression results, thus removing variation from the error term. Doing this, we make the estimate of β less biased as the variation is now included in the regression rather than lumped together in the error term.

$$Y_i = \alpha + \beta Cov_i + \theta X_i + \mu M'_i + \varepsilon_i \quad (2)$$

Above we see an example of adding control variables. We can recognize most of the regression from earlier, but we have now added two more parts. We add X as a control variable in order to control for individual variation. It is important to remember that while introducing X adds a part to the equation, it does not actually add more variation or possibilities to the regression. Rather, it only isolates a specific variation and sets it apart, allowing us to measure its effect. The effect is again measured by estimating the accompanied θ, as X takes the value of the individual. In this analysis, X will include gender, whether the individual was born in Norway, number of siblings, parental income (measured before the child turns 7) and cohort fixed effects. These variables will be individual terms and have an individual estimated effect. We include these variables in this manner both to remove the variation from the main estimate, and to allow us to see the estimated effect they have on the individuals in the analysis.

In addition to the X, we have added the term M'. This is an example of multiple control variables being grouped together and it can serve a few purposes. It is useful when we wish to control for variation between individuals which cannot easily be presented through variables. In this analysis, we need to control for fixed effects of the municipality the individual lived in at the age of 13. As there can be major differences between municipalities, controlling for fixed effects within each municipality is pivotal to avoid causing bias to our estimate, due to covariation amongst individuals from the same municipality.

5.2 The DID assumptions

When we control for municipality fixed effects, we are in practice using a difference-in-differences model. The basis of this model is to compare changes in two areas, where only one of them has been affected by the exogenous change, in our case the expansion. By using municipalities affected by the expansion as a treatment group, and those unaffected by the expansion as a control group, we can work on finding a causal relationship between Internet access and educational attainment. This is the case because the control group will also be affected by other changes affecting the treatment group, allowing us to control for changes affecting the results which is not the expansion. Without the use of the control group, the estimate will be affected by other variations than the expansion, such as an educational reform.

The main assumption we need fulfilled to be able to use this model is the assumption of common trend, meaning the two groups should have the same trend before the expansion hit. Optimally, we would compare the outcome of the same individuals and their outcomes with and without the expansion. As this is not possible, as an individual can only be affected, or not be affected, the next best thing is to use a contrafactual outcome. This is where we use the control group, which in a perfect scenario consists of individuals identical to those affected, the only difference being that they were not affected. If this is the case, they will have a common trend before the expansion, and the difference in trend after the expansion will be the effect of the change.

For this reason, the choice of control group is very important to this approach. We have to be able to argue the outcome of the two groups would be the same if the expansion had not taken place. While most Norwegians are fairly alike compared to someone from a different part of the world, there are definitively differences between individuals growing up in Oslo compared to rural Northern Norway. These differences can cause the estimates from the analysis to be inaccurate, and our results not being a good reflection of reality.

All we need for a basic version of the model is four data points. We need values of the outcomes before the expansion for both groups, and values for the outcomes after the expansion for both groups. While the difference between before and after for the treatment group will include the expansion and everything else that changed in the period, the difference between before and after of the control group will only include changes other than the expansion, as they were not affected by the expansion. By taking the difference between the two differences, we can find the change caused by the expansion while removing the change caused by other factors. This also removes fixed effects, as the starting and ending values are compared within the group first, thus removing the fixed effects.

$$\delta_{DD} = \left(Y_{1,1} - Y_{1,0}\right) - \left(Y_{2,1} - Y_{2,0}\right) \quad (3)$$

We can set the basic model up like the equation above, where β is the estimated effect and Y is the outcome, and compare two areas in period 0, before the expansion, and in period 1, after the expansion. The two terms find the difference in the two periods within each area, and by removing the differences in area 2 from the differences in area 1, the remaining difference will be that caused by the expansion, assuming that is the only difference between the two areas.

The model can also be set up in a regression equation, which would give us estimates for the effect of the treatment, being the expansion, and the effect of the change in time period (Angrist and Pischke, 2014). Additionally, this makes it easy to introduce control variables.

$$Y_{dt} = \delta_0 + \delta_1\, TREAT_d + \delta_2\, POST_t + \delta_{DD}\, (TREAT * POST)_{dt} + \delta_3 X_{dt} + \delta_4 M'_{dt} + \omega_{dt} \quad (4)$$

Where Y is the outcome, and the subtext of d is equal to one if affected by the expansion and zero if not, and t is the time period. β estimates the effect of being treated, γ estimates the effect of time, and δ estimates the effect of both being treated and the effect of time. The X term is similar to the term in OLS, in order to allow for control variables, and ε is the error term.

Equation 4 is fairly similar to equation 2, the major difference being the number of terms we use to explain the effect. While we in OLS explained the whole effect solely using the coverage term, the effect is now explained through three terms. These include the effect of getting the treatment, in the form of Internet access, the effect of time moving from before to after the expansion, and the interaction between the two. For an individual who was not affected by the expansion, the treatment would be equal to zero, leaving only the time-variable to be different from zero. An individual who was affected will have an estimated effect from each of the three, and the difference between the two will therefore remove the effect of time and give us an estimated effect of the treatment.

When we are estimating the regressions, we control for fixed effects of both time and municipality. As post estimate the effect of time changing, the inclusion of controls for cohort effects will remove most of the variation. However, as the expansion takes place over multiple years, there could still be some difference between individuals before and after the expansion. As for place, the treatment is decided on municipality level, so it will be the same for all individuals from the same municipality. Controlling for municipality fixed effects will not affect the estimate as it is showing the difference between treated and not treated. It can however help reduce the standard error as it controls for variation within the groups of treated and not treated.

5.3 Heterogeneity of Impacts

Next, we will introduce an interaction variable. Previously, the possibility of gender differences was discussed but only to the extent of controlling for it, to make our estimate less biased. Including the variable for gender solely as a control variable allows us to remove this variation from our main estimate. While this does let us see the difference in educational attainment between genders and control for it, it does not show us whether the genders are affected differently. To allow this in the regression, we will introduce an interaction variable between gender and Internet access. The interaction variable will multiply the gender variable and the Internet access variable, and therefore be equal to zero if either of these are zero. As male is equal to one, the interaction will only take a value different from zero for male youth. In addition, it will scale with Internet coverage, and it cannot take a value above one as neither of the variables can be more than one.

$$Y_i = \alpha + \beta Cov_i + \mu(cov_i * \ male_i) + \theta X_i + \mu M'_i + \varepsilon_i \ (5)$$

The estimate of the interaction variable will therefore measure the effect of increased Internet access on male youth, which does not affect female youth. It also works as a control variable in the regression, making our Internet access estimate no longer include gender-specific effects and providing us a better estimate of the effect on all youth. The new regression equation will now include the interaction variable by using multiplicative, with the estimate outside the parenthesis as it is the estimate for the interaction rather than one of the variables.

5.4 Mother fixed effects

Difference-in-differences allow us to control for big changes other than the one we wish to estimate the effect of. However, as we do not have a perfect control group, we still need to control for variation in individuals. Similar to OLS, controlling for unobservable variation is difficult but it is necessary in the attempt to isolate the effect of Internet access. An empirical approach we can use to attempt to solve this is based on using the shared effects of a mother on their children. This approach is called mother fixed effects, where I use the fact that siblings share a lot of characteristics not visible in the data. By comparing individuals who share mother, I can control for an important fixed effect covering socio-economic status, parental background and more. This will both allow me to control for more observable variables, rather than attempting to obtain data on everything, as well as make it possible to control for some unobservable variation not possible to control for using OLS. All this assists me in isolating the effect of the change in 4G-coverage, making the estimate from the regression less biased than the OLS-estimate.

To use this approach, we need to use data on mothers and their children in order to compare siblings. While using this approach is only possible if we look at siblings, I will still use data on all individuals, both with and without siblings. This is important, as if I were to only look at mothers with at least two children, it would allow for effects of having siblings and other effects coming from having bigger families. This would create an additional obstacle when attempting to isolate the effect of a change in an observable variable, and introduce new bias to our estimate, in our attempt to make it unbiased. The regression will look like this:

$$Y_{im} = \alpha + \beta Cov_{im} + \theta X_{im} + \delta M'_{im} + F_m + \varepsilon_{im} \quad (6)$$

Where Y is the educational outcome, Cov is the coverage at the age of 13, X and M' are control variables, F is mother fixed effect, and ε is the error term. While most of the terms can be recognized from the OLS-regression, there are two major differences. First of all, the introduction of a term for mother fixed effect, allowing us to control for the effects of the mother in order to compare siblings. The other is the introduction of small m's to accompany the i's. As we know from earlier, the i indicates that the variable takes the value of the individual. This is still true, but now we also have the m, which signifies the mother. While this does not affect the value of the variable, as it is still the value of the individual we use, it ties siblings together. We can also note the lack of an i on the mother fixed-term, as this shows the variable will take the value of the mother, rather than the individual, and therefore be shared amongst siblings. When we do compare siblings, finding the differences between them will remove the effects of the mother, as it is fixed.

6. Results

In this section we will go through the results of the regressions in the analysis. We will present the results of different regressions we have run, gradually introducing new elements, and explaining the reason why they are included and what effect these have. First, we will go through the OLS-model thoroughly, establishing a good overview of the situation we are looking at. Next, we will present the results of the mother fixed effect-model, comparing the results to the results of the OLS-model and discussing the differences of the two.

6.1 Main results

In the analysis, we will be looking at the three different measurements of educational attainment previously mentioned. We will look at the three different outcomes, in three different tables, in the order of GPA in middle school, followed by GPA in high school and finally high school completion rate. The three tables will gradually introduce new control variables in each column, starting with some individual characteristics and followed by some parental characteristics. For each column there will be control variables introduced, and they will be kept for the rest of the analysis. Therefore, a control variable introduced in column 1 will still be there in column 2, even though the focus will be the change caused by the newly introduced variable. The outcome variable used in a specific regression will be marked on top of each column, while the independent variable and the control variables will be marked on the side.

It is important to remember that the estimates of the control variables are not estimating the effect of change in Internet access. The estimates of our control variables show the correlation of the variable has on the measure of educational attainment, and in doing so make our main estimate, Internet access' effect on educational attainment, less biased. We do this to control for the aforementioned omitted variable bias. Additionally, if an estimate is marked with stars, it signifies that the estimate is statistically significantly different from zero at different significance levels. If an estimate is marked with a single star, it is statistically significant on a 10%-level, but not on a 5%- or 1%-level. The levels are based on how many percent of the observations in the sample are expected to be within the range. If an estimate does not have any stars, we cannot assume it to be different from zero. This is based on the size of the estimate and the standard error.

The standard error signifies how much variation is expected, based on how much each individual varies from the mean, the mean being our estimate. We have clustered the standard error on municipalities. By clustering on municipality-level, we account for covariation of the standard deviation amongst individuals from the same municipality. Clustering will typically increase the standard error, assuming individuals in clusters positively covariates. This is the case in our analysis, and it will lead to the need for higher estimates in order for them to be statistically significant.

Middle school GPA

Table 2: Dependent variable: Middle school GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cov	0.093***	0.025	0.028*	0.028*	0.016	0.016	-0.000
	(0.023)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)
1 [Male]			-0.411***	-0.411***	-0.413***	-0.413***	-0.416***
			(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
1 [Born in Norway]			0.163***	0.161***	0.051	0.021*	0.056***
			(0.039)	(0.037)	(0.033)	(0.012)	(0.010)
# siblings				-0.014**	-0.003	-0.003	-0.015***
				(0.006)	(0.006)	(0.005)	(0.002)
1 [High income]					0.344***	0.340***	0.125***
					(0.010)	(0.009)	(0.005)
Municip. FE	No	Yes	Yes	Yes	Yes	Yes	Yes
N	224520	224520	224520	224520	216417	214983	213142

*Note: Estimated effects are to show the difference between an individual who had access to 4G-Internet and an individual who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, regressions (6) and (7) includes whether parents are born in Norway, and regression (7) includes educational level of parents. *p<0.10, **p<0.05, ***p<0.01.*

Table 2 includes the estimates of the regression using GPA in middle school as the outcome variable. In column 1 we have only included control variables for birth year fixed effects, and the estimates of these are not included in the table. It is important to control for year-by-year changes due to long-term trends, as it can give the individuals a different basis depending on their year of birth. However, it is not something an individual can change, or something that can be changed at all. The data could tell us that those born in 1994 have better educational attainment than those born in 1993, and while we should control for this for the sake of our main estimate, there is nothing else we can use it for.

Regarding the estimate of interest, coverage, which we can see is 0.093 and is marked with three stars. While coverage in our data is on a scale from zero to one depending on the coverage rate in the municipality, for the effect on an individual we will consider it as either zero or one, as an individual either has access to the LTE-network or they do not. Considering we have only controlled for the fixed effect of birth year, there are many other factors that likely are at play but are still in the error term, making our estimate biased. Regardless, our regression estimates that an individual who was given access to the LTE-network would increase their grades by around 0.1, which is fair bit lower than the standard deviation of middle school GPA we found earlier of 0.8. While a change from 4.1 to 4.2 is not too significant, our main takeaway from this regression is that the effect is positive.

Before introducing more control variables we want to take potential within group variance into account. Within group variance can occur if individuals in the sample are grouped up, and the variation of an individual is affected by difference in the groups. In our case, this is likely as the individuals are from different municipalities and this will likely affect them. Municipalities have different quality of education, and individuals growing up in a municipality with high-quality education will do better than those growing up in a municipality with lower quality. This is however not something we want to include in our estimate, as we want to find the effect of Internet access regardless of the quality of education. To do this, we want to control for municipality fixed effects in column 2. This will absorb all permanent differences in municipalities which both affects educational attainment and correlates with the Internet coverage. After including this in our regression, our estimate falls drastically and while still positive, it is no longer statistically significant. In addition, the estimated standard error has risen slightly.

Then, we can start introducing the control variables. A natural start is introducing gender, and we can see our estimate is clearly negative. This estimate shows us boys on average get lower grades than girls in middle school. In addition, we control for whether the individual is born in Norway or not, as difficulty integrating into Norway can have an effect on educational attainment. This estimate is positive, meaning Norwegian-born youth has better grades than non-Norwegian-born in middle school. More importantly for this analysis is how the inclusion of these control variables affect the main estimate, which we notice it has done. The estimate is now a bit more positive, leading to it being statistically significant on a 10% significance level.

Next, we control for number of siblings in column 4, as family size could have an effect. As the estimate is negative and statistically significant, it tells us that having siblings can have a minor negative effect on middle school grades. In addition, unlike the previously included control variables, this variable is not a dummy variable, so it is not bound to be either zero or one. While we could have included it as a dummy showing whether the individual has siblings or not, we have chosen number of siblings as the control variable. Therefore, if an individual has multiple siblings, our estimate tells us this would have a larger negative effect than just one sibling, as it shows the effect of increasing the number of siblings by one. It is however not possible for this model to give a different effect of increasing from, for example zero to one or from one to two, so if there is a difference the estimate will be a weighted average of the effects. The effect of adding it on the other estimates in the regression is rather small, as the rest are near identical to column 3.

In column 5 we have added the first parental characteristic, being high income. As explained earlier, this variable is equal to one only if both parents of the individual earn more than the average income in Norway. The estimate for this variable is very positive and statistically significant, so individuals with high-earning parents tend to do better than those without. Controlling for parental income also has a significant impact on three of the other four included estimates. Mainly, it reduced the Internet access estimate a fair bit, so while it is still positive, it is no longer statistically significant. Additionally, while the estimate of the effect of gender stays nearly the same, estimates for both being born in Norway and number of siblings are no longer statistically significant. This does not mean that we should exclude them as control variables, but it does tell us that the previous estimates for these were likely partly because of difference in parental income. As the parental income was not included as its own control variable previously, the data explained some of the correlation with high-income parents through the other control variables. Practically, this means the effect of being born in Norway is not as large as the regression in column 4 estimated, as part of the effect came from Norwegian-born individuals more often having high-income parents than foreign-born individuals.

In column 6 we introduce a control for whether the parents of the individual are born in Norway, similarly to what we do for the individual themselves. This variable is a dummy, but unlike high-income, both parents have a variable assigned, which allows for one to be equal to one and the other being equal to zero. While we have not included the estimates in our table, we can see it changes the estimate of the effect of an individual being born in Norway. We see it brings the estimate closer to zero, but it also brings the standard error much lower, resulting in the estimate again being statistically significant on a 10%-level. The new estimate still tells us the effect is positive, but it is smaller than earlier.

Lastly, we include parental education in column 7. This brings the estimate down to zero, telling us that the regression estimates no connection between Internet access and grades in middle school. The control variable estimates change some as well, but the direction stays the same. We do see that controlling for parental education lowers the standard errors, leading to all the control variable estimates being statistically significant at a 1%-level. This tells us that parental education is quite relevant, explaining a lot of the variation in other control variables included.

What does this mean for our main estimate? The estimate for Internet access varies a fair bit, depending on which control variables were included. Up until we included parental education in column 7, the estimate was positive, and some of the time statistically significant. While this gives us an inclination to assume the effect is positive, it is important to remember the last regression also controls for all the same variables as the previous regressions. Therefore, as the last regression controls for most relevant variation, it could be the least biased estimate. However, the inclusion of controls for specific parental variation could also affect the estimate for the worse as we have already controlled for a lot of the variation through whether the individual is born in Norway, introduced in column 3, and the parental income, introduced in column 5.

Our main finding from this part of the analysis is therefore that there is no relationship between access to the LTE-network and grades in middle school. Inclusion of control variables is not guaranteed to make the estimate less biased, as there are bad control variables. The use of bad controls would cause the estimate to be more biased as they also are dependent on the independent variable. This could lead to the estimated effect being biased through the effect of the independent variable on the bad control variable.

High school GPA

However, as discussed earlier, we need a wider range of outcomes if we wish to look at the effect on educational attainment. For now, we have only looked at grades in middle school, giving us a look into the effect on youth at the ages of thirteen to fifteen years old. Going into high school, youth start having more freedom and opportunities to choose themselves. Does the added freedom of the Internet have a positive or negative impact on their educational attainment in their high school years?

We want to run this analysis similarly to the previous one, as this will give us the opportunity to easily compare them. In the table below we can see the estimates for high school grades. The only change in this analysis compared to the last one is that our outcome has changed from grades in middle school to grades in high school. These are based on the same scale, one to six, and while high school grades do allow an average grade slightly above six due to bonus points stemming from high-difficulty courses, the difference is not large enough to mention further. All control variables included are the same, and they have been introduced at the same time as previously.

Table 3: Dependent variable: GPA at the end of high school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cov	0.102***	0.015	0.017	0.017	0.004	0.005	-0.007
	(0.024)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
1 [Male]			-0.283***	-0.282***	-0.282***	-0.283***	-0.285***
			(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1 [Born in Norway]			0.209***	0.208***	0.122***	0.055***	0.088***
			(0.033)	(0.032)	(0.028)	(0.011)	(0.010)
# siblings				-0.008	0.002	0.003	-0.008**
				(0.006)	(0.006)	(0.006)	(0.003)
1 [High income]					0.263***	0.258***	0.092***
					(0.013)	(0.011)	(0.007)
Municip. FE	No	Yes	Yes	Yes	Yes	Yes	Yes
N	221511	221511	221511	221511	213634	212250	210553

*Note: Estimated effects are to show the difference between an individual who had access to 4G-Internet and an individual who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, regressions (6) and (7) includes whether parents are born in Norway, and regression (7) includes educational level of parents. *p<0.10, **p<0.05, ***p<0.01.*

The estimates for coverage in Table 3 are similar to those in Table 2. In column 1, we again get a strong positive estimate, statistically significant on the 1%-level. Then, when introducing absorption to control for municipality differences, the estimate falls drastically and is no longer statistically different from zero. This tells us that while there are likely differences in effect, a lot of it can be explained by the differences between the areas individuals grew up in.

Continuing onto the inclusion of individual characteristics, we should note one interesting difference. The Internet access estimate is not statistically significant after controlling for individual characteristics, like it was in the middle school estimate. The variation is nearly the same, but the estimate of the effect is a fair bit lower. We can also see there are some differences in the estimates of the control variables. While young men do worse also in high school compared to their female counterparts, the effect is estimated to be a bit lower now than it was in middle school. The estimate for Norwegian-born is still positive, but a bit higher than previously. Lastly, while the effect of having siblings is still negative, it is not statistically significant in here, unlike the middle school estimate.

When we introduce parental characteristics, we again see the Internet access estimate is lower than it was previously. Our middle school estimates were around one standard error away from the mean, which while not statistically significant gives an inclination to believe the effect is at least not negative. Now however, it is much closer to zero and when we control for parental education, the estimate turns negative. Similarly to the last table, column 7 gives the least biased estimate as we have included the most control variables. As it is still low and not statistically significant, we do not assume the effect to be negative, but it is the first indication of the effect not being positive.

High school completion

The final outcome variable we will look at is high school completion at the age of 19. While we will still use the same control variables, and introduce them in the same fashion, the estimates will likely be a bit different. Both middle school and high school grades are on the standard Norwegian scale for grades, while high school completion will be a dummy variable, equal to either zero or one depending on if the individual had completed high school when they were 19 years old. Our estimates will therefore not show the effect on the outcome in the same way as earlier, but rather the effect it has on the probability of an individual completing high school by the age of 19. However, it is still a measure of educational attainment, so we could expect the effect of Internet access to be in the same direction as with grades. If it has a positive effect on grades and a negative effect on high school completion, there would likely be some underlying differences which we have not accounted for. With that in mind, we can look at the estimates in Table 4.

Table 4: Dependent variable: High school completion at the age of 19

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cov	0.115***	0.023**	0.025**	0.025**	0.019*	0.019**	0.012
	(0.020)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
1 [Male]			-0.196***	-0.196***	-0.197***	-0.198***	-0.199***
			(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1 [Born in Norway]			0.069***	0.068***	0.018*	0.022**	0.038***
			(0.013)	(0.012)	(0.010)	(0.007)	(0.006)
# siblings				-0.006***	-0.001	-0.001	-0.007***
				(0.002)	(0.002)	(0.002)	(0.002)
1 [High income]					0.160***	0.159***	0.070***
					(0.004)	(0.004)	(0.003)
Municip. FE	No	Yes	Yes	Yes	Yes	Yes	Yes
N	201017	201017	201017	201017	193775	192512	190860

*Note: Estimated effects are to show the difference between an individual who had access to 4G-Internet and an individual who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, regressions (6) and (7) includes whether parents are born in Norway, and regression (7) includes educational level of parents. *p<0.10, **p<0.05, ***p<0.01.*

Before introducing any control variables, the estimates again show a positive effect, which is heavily lowered when we remove municipality-differences. However, unlike earlier, the estimate is still statistically significant on a 5%-level. This trend also continues when we control for individual characteristics, and we can see the Internet access estimate actually increases slightly when controlling for this. We also recognize the directions of the control variable estimates, though they are naturally a bit lower than when the outcome is grades as previously mentioned, the estimates all go in the same direction as they did with the previous two outcome variables. Column 5, which includes all individual and family characteristics, estimates the effect to be positive, and statistically significant on a 10%-level. The inclusion of parental income also brings down the estimate for the individual being Norwegian born, showing the effect was likely driven by the lower average income of non-Norwegian born.

At last, introducing parental characteristics does not bring any major changes. Like with the previous outcomes, the Internet access estimate is lowered further, and is no longer statistically significant. As for the control variables, the estimate for gender stays nearly identical regardless of what other control variables are included. The estimate for Norwegian-born becomes less statistically significant when only introducing high-income parents, but the difference seems to be explained by place of birth and educational level of the parents, as when these control variables are introduced the estimate goes back to being positive and statistically significant on a 1%-level. We also see that the introduction of parental education makes the negative estimated effect of siblings statistically significant and cause the estimated effect of high-income parents to be lower, but still clearly statistically significant.

Important to note is that we should be careful reading too much into specific estimates when using OLS to look at probability, such as when the outcome is a dummy variable. As the OLS-model estimates a linear relationship between the outcome and the independent variable, it does not limit the outcome to being between zero and one. This could cause the predicted outcome of the outcome variable to be less than zero or more than one, which is not possible as probability only makes sense between zero and one hundred per cent.

All three of the chosen outcome variables are positive outcomes of educational attainment. Youth having better grades in middle school and high school, and higher high school completion rate, are all positive outcomes which will likely be affected by similar effects. For this reason, all three of our regressions have shown quite similar effects. While this is expected, we can still not claim to have found a causal effect.

Based on the results presented in Tables 2, 3 and 4, we will choose the specification in column 5 as our main regression. We do this as the addition of more control variables does not seem to affect the estimate noticeably, as most of the variation is already controlled for through the parental income control variable. Additionally, we will drop the sibling control variable, as it does not affect the estimate much, and we will be using a mother fixed effects approach later which does not need the control for siblings. This leaves us with our preferred set of control variables being gender, Norwegian born, high-income parents, and year-by-year effects.

6.2 Heterogeneity of impacts

A very interesting and relevant effect we should allow for is the possibility of Internet access affecting young men and women differently. While we already have controlled for gender, and it that way taken it out of the equation, we have not allowed for the effect to be different on the two genders. By including the interaction variable proposed in the empirical strategy-chapter, we can allow our model to estimate a different effect based on gender. When including the interaction variable, Internet coverage will be part of two variables. As we are estimating the effect of coverage, estimating the effect of both individually is not enough. We also have to test whether they can both be equal to zero at the same time. The result of this test is presented in the bottom of the table, where p-value shows the probability of both coverage-variables being equal to zero at the same time.

Table 5: Gender-specific effects of increased Internet access

	(1) MS GPA	(2) HS GPA	(3) HS-completion by age 19	(4) Any Child before age 20
Cov	0.033**	0.027	-0.002	-0.004**
	(0.016)	(0.017)	(0.010)	(0.002)
Cov X Male	-0.034***	-0.045***	0.039***	0.015***
	(0.009)	(0.012)	(0.012)	(0.001)
1 [Male]	-0.408***	-0.275***	-0.199***	-0.015***
	(0.007)	(0.005)	(0.005)	(0.001)
p-value	0.0006	0.0008	0.0014	0.0000
N	216417	213634	193775	267241

*Note: Estimated effects are to show the difference between an individual who had access to 4G-Internet and an individual who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, Norwegian born dummy, and high-income parents dummy, and control for municipality fixed effects. *p<0.10, **p<0.05, ***p<0.01. P-value shows the probability of the estimates of both coverage-variables being equal to zero.*

Gender

Before introducing the interaction variable, our estimate inclined us to believe there is a positive relation between Internet access and educational attainment, regardless of which outcome we looked at. However, after introducing all control variables, there were no statistically significant effects and the estimate of effect on high school grades actually turned slightly negative.

In table 5, we have now introduced the interaction between Internet access and young men. To avoid further repeating explanations on different control variables, we have now chosen only one combination of variables, and presented all the three outcomes using this combination. This combination is nearly the same as column 5 in the previous tables, except the removal of siblings as a control variable. While our regressions showed siblings could have a slight effect, we have decided to remove it as the effect was very low. Remaining, as shown in the table, we have included gender, Norwegian-born and high-income parents as control variables, in addition to the interaction variable. We have controlled for municipality-effects and year-by-year effects in the same way as earlier.

The estimates are now telling a bit of a different story than earlier, and the estimates for each of the educational attainment outcomes are also different. In column 1, we see the estimated effects on middle school grades. Comparing these to the previous estimates, in Table 2 column 5, the estimated effect of Internet access is higher and now statistically significant. However, we find an equally negative estimate on the interaction between male and Internet access. This tells us there is likely a gender-specific effect, while girl's grades are positively correlated with Internet access, boys do not seem to experience the same effect. As for the control variables, the estimates are very similar to the estimates from the regression in column 5 from earlier. The regression still estimates boys to have lower grades than girls, even when accounting for the seemingly different effects of Internet access. The estimated effect of being born in Norway is still positive but not statistically significant, and high-income parents have a clear positive effect on middle school grades.

Onto high school grades, we can see the estimates are fairly similar to the middle school estimates. First, the estimate for Internet access is somewhat smaller, and this causes it to no longer be statistically significant on a 5%-level. The estimate of effect in specifically young men is somewhat higher but so is the estimated standard error, making the total difference rather small. The estimates of the control variables are again very close to estimates we got before adding the interaction variable, young men perform worse than their female counterparts, Norwegian-born youth gets higher grades than those not born in Norway, and having high-income parents correlates with better grades in high school as well.

When it comes to high school completion, our estimates now look a bit different than before. Without the interaction variable, there was a statistically significant positive effect of Internet access, and now, the estimate is approximately to zero. Even more interestingly, the estimated effect of the interaction is clearly positive, unlike the estimated effect on grades. While it still looks like young women have a higher probability of completing high school within the age of 19, the addition of Internet access seems to have a positive effect on male completion rates. Earlier, we mentioned the expectation of the effects being in the same direction as higher grades typically correlates with higher probability of completing high school. However, this introduces the possibility of Internet access helping some young men stay in school, even if the effect on the grades for all young men is estimated to be negative. The addition of the interaction variable did not change the estimates of the control variables, these are still estimated to have the same effect as the regression without interaction, and the regressions using the other outcomes.

Lastly, to get a different view of the situation by changing the dependent variable to something based on behavior. A behavioral outcome could shed some new light, while still be relevant as it could also affect educational attainment. For this, we will look at teenage pregnancies, as this is a massive change in an individual's life and will likely affect their educational attainment (Levine and Painter, 2003). This is not whether the mother gave birth as a teenager, rather, it is now the individual themselves.

To do this, we make a dummy variable equal to one if the individual had their first child at the age of 19 or earlier, otherwise equal to zero. The gender of the individual does not matter, the variable will be assigned to both men and women who were a teenager at the time of birth of their first child.

We are running this analysis on behavior to look at the same situation from another angle, and we will therefore use this dummy variable as the outcome. While we earlier used the three different outcomes of educational attainment, we will now use this behavior as outcome, and keep everything else the same. This will ensure that we are estimating the effects of a similar situation, even though the outcome is something distinctly different. Something to keep in mind is the low number of teenagers becoming parents, leading to the estimated effects being quite small. However, this will also cause the standard error to be small, meaning the estimates can still be statistically significant. We should also keep in mind that as the outcome variable is a dummy, the regression will give estimates on change in probability in the same manner as when we looked at high school completion rates.

Interestingly, we do get a statistically significant estimate, signifying a negative correlation between Internet access at 13 and having a child before turning 20. Additionally, the effect is estimated to be quite different based on gender. It shows us that young women have a higher probability of having a child before they turn 20, compared to young men. However, young men are affected positively by the Internet access, unlike young women.

Parental Education

The interaction between gender and Internet access is not the only interaction we have an interest in looking at. While we initially controlled for parental education, we did this by allowing each level of education to have an individual effect. Because of the low differences between each level, the effect will likely be small and might be affected by other factors.

However, if we can find a natural divide which we can use to reduce the number of groups, we can attempt to use this to look at the effects. A difference in educational level which we can use for this purpose is whether or not a person has a degree from higher education. We will take advantage of this, creating a dummy variable equal to one if either one or both of the individual's parents have at least a bachelor's degree.

Furthermore, we use this dummy to create the interaction variable, combining it with the Internet access variable. Then, we will run the regressions in a similar way as the previous one. The slight difference will be that we still include gender as a control variable even though we do not have the gender interaction variable. Additionally, we now have a control variable for whether the individual's parent has a degree, as well as the interaction between degree and Internet access. Effectively, we will look at whether having highly educated parents influences how the increase in Internet access affects an individual.

Table 6: Effects by parental education

	(1) MS GPA	(2) HS GPA	(3) HS-completion by age 19	(4) Any child before age 20
Cov	0.020	-0.007	0.025**	0.000
	(0.015)	(0.016)	(0.010)	(0.001)
Cov X PC	-0.076***	-0.015	-0.043***	0.012***
	(0.009)	(0.011)	(0.008)	(0.001)
1 [Parent college]	0.579***	0.465***	0.236***	-0.010***
	(0.009)	(0.009)	(0.006)	(0.001)
p-value	0.0000	0.3466	0.0000	0.0000
N	215606	212829	193048	265864

*Note: Estimated effects are to show the difference between an individual who had access to 4G-Internet and an individual who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, Norwegian born dummy, and high-income parents dummy, and control for municipality fixed effects. *p<0.10, **p<0.05, ***p<0.01. P-value shows the probability of the estimates of both coverage-variables being equal to zero.*

In column 1 we see the estimates using middle school grades as the outcome variable. We are still including the same control variables as the last interaction variable regression, which was similar to column 5, but we have now also included parental education. While it was controlled for in a different manner earlier, it makes this regression similar to column 7 as well. First, we can see the Internet access estimate for those whose parents do not have a college degree is positive but not statistically significant, which is between our initial estimates and the estimate where we included the gender interaction. Like we found when we controlled for parental education initially, it seems like it accounts for some of the differences between Norwegian and non-Norwegian born individuals. As for the newly introduced variables, we see something interesting. While having highly educated parents show a strong positive correlation on the individual's grades in middle school, there is a reduction in the grades of children with higher educated parents in areas with high Internet coverage. Both estimates are also clearly statistically significant.

For high school grades in column 2, the Internet access estimate is negative, and not statistically significant. We saw a similar estimate in column 7 in the initial regression, making it seem like Internet access does not necessarily have a positive effect on high school grades. As this change is caused by controlling for parental education, it could mean the estimate is skewed by the positive effect of higher educated parents. Further, the story of the other control variables is similar to that which we just discussed, the estimates are in between column 5 and column 7 in the initial regressions.

The estimate for highly educated parents is again strongly positive, though it is a bit lower than what it was for middle school grades. The interaction variable has been estimated to be negative, like in the regression using middle school grades, but the estimate is not statistically significant. The fact that neither estimate involving Internet access is statistically significant in this regression gives us little to work with for interpretation. This is also reflected in the p-value, which is now nearly 0.35, which tells us there is around 35% probability that both estimates are equal to zero.

On to column 3, showing the probability of high school completion. This regression gives us a statistically significant estimate for the effect of Internet access, showing that individuals with lower educated parents look to be positively affected by the Internet. The estimate is also higher than both column 5 and column 7 in the initial regressions in Table 4, now being significant on a 5%-level. Furthermore, all the other estimates are also statistically significant in this regression. Gender, Norwegian born and high-income parents give the expected effects, so we will focus on the newly introduced variables. Just as with grades in middle school and high school, having highly educated parents seems to have a strong, positive effect on the individual completing high school. So, our estimates show it does not necessarily have a negative effect on high school grades, but it does seem to have fairly strong negative effect on completion rates.

The p-value also reflects this, as it finds the probability of both Internet access variables to be zero nearly impossible.

In column 4, we have again included the behavioral outcome of having a child before turning 20 years old. The estimated effect of Internet access is now zero, unlike the negative estimate we got earlier. We also see a strong negative effect of having highly educated parents. However, the interaction is clearly positive, telling us the effect of having highly educated parents seems to be nulled out by the increased Internet access.

In addition to these results, where we include an interaction between Internet access and gender or parental education, there is something important to note. The effect can be vastly different on different groups. While not unsurprising, this is very important to keep in mind if we were to discuss changes to be made based on the results of the analysis.

6.3 Assessment of validity and robustness

Until now, we have used control variables in an effort to isolate the effect of Internet access, making our estimate as unbiased as possible. While the control variables do make our estimate less biased, it is very difficult to say whether it is actually unbiased or just a bit less biased. If the latter is true, we cannot use the estimate to say something about reality regardless of the help contributed by the control variables, as the estimate is still biased. There is no perfect way to test this, but in this section, we will attempt to show that the regressions are well specified and have given us unbiased estimates. As mentioned earlier, we decided on the set of control variables originally introduced in column 5 of Table 2-4, as we believe this combination is well specified and will therefore give the least biased estimates.

For the estimates to be unbiased, the model needs to fulfill the identifying assumption of DID, which is the assumption of common trend. As we are using the unaffected individuals as a contrafactual value for the affected individuals, they need to have a common trend in the time before the expansion. We need to test this assumption in order to support the claim of unbiased estimates. This can be done using a multitude of different tests.

The first method we will use is running a placebo test (Hartman and Hindalgo, 2018). In medical research, the patients can be supplied with either the treatment, which is being tested, or what they believe is the treatment, but which in reality is not. As the patients do not know whether they got the treatment or the placebo “treatment” the difference in outcome will be the effect caused by the treatment as long as prior differences are accounted for. In econometrics, we cannot do the test in such a direct manner, as we cannot give some a treatment and some not, and we typically cannot access a perfect experimental setting where everything else is equal.

However, we can imitate the placebo test by running the regressions again, but using variables which should not be affected by the expansion of Internet during early adolescence. By testing something where there should be no effect, we will be able to either confirm the lack of effect or find an underlying difference which we had not accounted for in our previous regressions. In this analysis, we can do a placebo test using variables that are determined way before the Internet access was ever relevant. To do this, we will be using birth outcomes, as there should be no correlation between birth variables and Internet access. For example, there should be no correlation between the birth weight of an individual and their Internet access at the age of 13. To get a broad view, we will use some birth variation specific to the individual and some specific to the mother and the pregnancy.

Table 7: Placebo-test

	(1) Birthweight	(2) C-section	(3) Pregnancy-length (in weeks)	(4) Mom married at birth	(5) Miscarriage
Cov	4.769	-0.005	0.080^{**}	0.003	0.007
	(11.284)	(0.009)	(0.040)	(0.009)	(0.043)
Mean	3535.255	0.156	39.477	0.531	0.714
N	255295	255492	240551	255492	255492

Note: Estimated effects are to show the correlation between Internet coverage and variables which should not correlate with coverage. Standard errors are clustered at municipality level. Regression based on 411 municipalities over 18 years. All regressions include year dummies, Norwegian born dummy, and high-income parents dummy, and control for municipality fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7 includes the estimates of the five different dependent variables used in the placebo-test. As this is a placebo-test, I will describe the estimates of coverage, not the control variables below. Starting at column 1, the dependent variable is birth weight. Here, we should keep in mind that since the variable is measured in grams, the variation is larger than we have seen in the previous regressions. However, this does not necessarily mean it has an effect, as the standard error is affected in the same manner. The estimate is just under five, while the standard error is over eleven, so the estimate is not statistically significant. This means we find no correlation between the birth weight of the individual and the Internet access they had at the age of 13. Additionally, the estimate is much smaller than the average weight. We see a very similar story in column 2, using whether the individual was born by c-section as the dependent variable. The regression estimates there to be no correlation between whether the individual was born by c-section and their Internet access at 13 either.

Column 3 does however find a statistically significant correlation, using the length of the pregnancy as the dependent variable. This estimate tells us there is a correlation between the length of the pregnancy and the Internet access of the individual being born when they are 13. While we do expect to find no connection, it is not unlikely that some of the tests will give a statistically significant result. In this case, it could be covariation stemming from both correlating with a third variable. As an example, we know the expansion hit central areas first, and these areas can likely offer better pregnancy follow-up than more rural areas, and it is easier for the mothers to get to a hospital when close to giving birth. Similarly, urban areas are typically richer, and more well-off mothers typically have longer pregnancies. We also see the change of 0.08 is very small compared to the mean of almost 40 weeks, making it only a 0.2% difference. Even though the estimate is statistically significant, it is a very marginal difference.

In column 4, we find no connection between the Internet access and whether the mother was married at the time of birth. In column 5, the dependent variable is whether the mother has had a miscarriage before. This is a dummy as to whether it happened or not, not a variable taking on the number of times. This estimate is very close to zero, showing that there seems to be no connection between Internet access and the probability of the mother having had a miscarriage.

The test shows us that most of the predetermined variables do not correlate with Internet access. This is the result we want, as it shows there is no underlying variation deciding both. As placebo-tests will give some statistically significant results by design if you test many variables, we view this result as an indication of no correlation.

6.4 Mother fixed effects

Using the mother fixed effects approach, we can control for the things such as family culture, which is not possible to acquire data on. We will run the analysis in a similar fashion to how it was done previously in this paper, but we will now compare siblings rather than comparing youth from the same municipality. While we are now using a different empirical approach, we still want our results to be based on similar grounds. Therefore, we will use the same combination of control variables as earlier. We have decided to use the control variables used in column 5 in Tables 2-4, similarly to what we did in Tables 5 and 6. This includes control variables for gender, Norwegian born, high-income parents, and year-by-year effects.

Table 8: Comparing siblings

	(1) MS GPA	(2) HS GPA	(3) HS-completion by age 19	(3) Any child before age 20
Cov	-0.035	-0.059**	0.007	0.002
	(0.027)	(0.030)	(0.020)	(0.003)
N	216417	213634	193775	267241

*Note: Estimated effects are to show the difference between a sibling who had access to 4G-Internet and a sibling who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies and controls for gender, Norwegian born and high-income parents. *p<0.10, **p<0.05, ***p<0.01.*

Table 8 includes the estimates for all three educational attainment outcomes. In column 1, we see the estimates for middle school grades. While the estimate is not statistically significant, it is negative. This goes against our previous findings, where the estimates were almost exclusively positive. The estimated effect of gender is still negative, meaning we still expect boys to have lower grades than girls. This ensures that differences between brothers and sisters are not wrongly accredited to the difference in Internet access. Similarly, being Norwegian born is still estimated to be positive, and including this in the analysis ensures differences between Norwegian-born and non-Norwegian born are not wrongly included. Interestingly, the estimate for high-income parents is no longer statistically significant, which it was in all the previous regressions we ran.

Column 2 uses high school grades as the dependent variable, and here the estimate is also negative, and additionally it is statistically significant. This makes it the first estimate of Internet access to be statistically significant negative estimate in our analysis and asks the question whether the OLS-regressions are missing something crucial, leaving us with misleading estimates. The rest of the estimates are similar to what we found in column 1.

Column 3 again shows us the estimated change in probability of completing high school on time. This estimate is not negative, rather it is very close to zero, seeing as the standard error is much higher than the estimate itself. The estimate of gender is still negative, showing it is still more likely for young men to drop out of high school, while both the estimate for Norwegian born and high-income parents are now just around zero.

This brings up an interesting difference. Not only do these estimates introduce the possibility of a negative effect on schooling grades, but it also estimates this difference to not affect whether the individuals actually complete high school on time.

As we did previously when using OLS, we want to allow for the model to estimate a gender-specific effect. To do this, we will use a similar regression as we did in Table 5, now using the mother fixed effect approach. We will use the same control variables as in Table 8, and additionally include the interaction variable introduced in Table 5. As a reminder, in Table 5 we found the estimated gender-specific effect to be negative on middle school- and high school grades, but to be positive on high school completion.

Table 9: Gender-specific effects when comparing siblings

	(1) MS GPA	(2) HS GPA	(3) HS-completion by age 19
Cov	-0.017	-0.045	-0.003
	(0.028)	(0.031)	(0.022)
Cov X Male	-0.036**	-0.028	0.019
	(0.017)	(0.021)	(0.023)
1 [Male]	-0.397***	-0.273***	-0.199***
	(0.014)	(0.009)	(0.010)
p-value	0.0580	0.0768	0.6739
N	216417	213634	193775

*Note: Estimated effects are to show the difference between a sibling who had access to 4G-Internet and a sibling who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, and controls for gender, Norwegian born and high-income parents. *p<0.10, **p<0.05, ***p<0.01. P-value shows the probability of the estimates of both coverage-variables being equal to zero.*

Table 9 includes the estimates using all three educational attainment outcomes as in Table 8, now with the interaction between gender and coverage included. In column 1, we find a negative but not statistically significant estimate like in Table 8. It is however smaller than previously, which is not hugely important for this estimate, but will be relevant in column 2. The estimate of gender is still clearly negative and of similar size. In addition, we have the interaction, estimating a further negative effect of the middle school grades of boys specifically. We have not included the estimates for the other control variables used as they were already discussed when included in Table 8, and they should not be affected by the introduction of the new variable.

The estimated effects on high school grades in column 2 have been affected similarly to those in column 1. The estimate for Internet access has been lowered and is now too low to be statistically significant. While the gender estimate is still statistically significantly negative, the new interaction is not, leaving both estimates for coverage to being not statistically significant. The estimates for high school completion in column 3 do not show any effect of coverage either, leaving only the gender estimate to be statistically significant. It is however worth mentioning that the gender-specific estimate is here positive, leaving us with results which resemble those we found in Table 5.

Lastly, we need to mention the p-values. As we can expect, the regression with a statistically significant estimate, in column 1, has the lowest probability of both estimates being zero. However, we can see the probability of both estimates being zero in column 2 is also rather low, much closer to the probability in column 1 than in column 3. This is because by the estimates, while not statistically significant, nearly being statistically significant as the estimates are higher than the standard error. This tells us that there is a negative effect for boys, as the probability of both being zero is very low. So, while neither column 2 nor column 3 have statistically significant estimates for the effect of Internet access, the regression in column 2 finds it likely that at least one of the two estimates are not equal to zero. Column 3 does not, as we can see there is a strong possibility that the two estimates are equal to zero at the same time.

While these estimates are consistently negative, and some are statistically significant, the concept of birth order effects. Research shows younger siblings typically take less education and have lower general educational attainment (Black et al., 2005). Taking this effect into account, we would need the estimates to be quite low to claim significance. As some only a few are statistically significantly different from zero it is possible that there is confounding between birth order effects and coverage expansion in this setup. More work would be needed to disentangle the two, but that is beyond the scope of this thesis.

6.5 Discrete Dependent Variables

There is also a problem with using OLS when the outcome variable is a dummy, as this will make the results estimate probability of the outcome happening. An OLS-regression finds a linear connection between the outcome and the independent variable, and the probability can be found by adding the values of an individual. However, it does not restrict the outcome to be between zero and one, as it should when we are using probability. Therefore, an individual could have lower than zero percent or higher than one hundred percent probability of the outcome, according to the model.

To avoid this, we need a model which restricts the outcome to be between zero and one. The solution is to use a maximum likelihood method, and we will be using probit. Using probit, we will get a non-linear estimate, allowing for individuals at different levels of probability to be affected differently, leaving all to have a probability between zero and one.

Until now, we have been using the OLS-model to run the regressions. In the process, we have introduced many different control variables in an effort to estimate the effect of Internet access, and it has allowed us to make a less biased estimate. Whether it is completely unbiased is difficult to know, it is likely that it is still biased, but we ran a placebo-test to further look at this.

However, there is a lot of variation we have not included in the analysis. While some variation can be measured and included, this is in many cases very difficult and time-consuming. As an example, measuring and including data on mental health would likely be inaccurate at best, as we would need to combine many different factors into one or a few data points. It could also introduce problems such as measurement errors, as the data could be misreported, or the individual might not even know the true answer themselves. In addition, some variation is not observable, making it almost impossible to gather it as data. Something such as family culture or ways of teaching discipline makes little sense to make into numbers.

Table 10: Marginal effects of probit model

	(1) HS-completion by age 19	(2) Any child before age 20
Cov	0.053**	-0.040
	(0.026)	(0.104)
N	193775	216821

*Note: Estimated effects are to show the difference between an individual who had access to 4G-Internet and an individual who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year and municipality dummies, and controls for gender, Norwegian born and high-income parents, and control for municipality fixed effects. *p<0.10, **p<0.05, ***p<0.01.*

Table 10 includes a maximum likelihood regression using the probit model for the two outcomes which are dummy variables, high school completion at 19 and having a child before 20. In column 1, the estimates show a statistically significant positive effect. This is a fair bit higher than what we found in Table 4 when we used a linear approach with OLS. However, the standard error is also higher, making both significant on a 5%-level. The estimates in column 2 show slight negative correlation, the variation is too high to claim there to be a between Internet access and having a child before turning 20. This is also consistent with our previous findings, as while we have not run the exact same model, the models with interaction variables estimated the effect to be negative, and not statistically significant respectively.

6.6 Coverage as a dummy variable

We can also convert the coverage variable to a dummy variable and use this rather than the continuous coverage variable we have been using. Doing this, the regression will estimate the effect of being given access to the high-speed Internet. Additionally, as the coverage variable does not reflect on the difference in Internet speed, whether an individual has coverage or not gives a good read of the situation.

Table 11: Coverage as a dummy variable

	(1) MS GPA	(2) HS GPA	(3) HS-completion by age 19	(4) Any child before age 20
1 [Cov>0]	0.016 (0.017)	0.017 (0.017)	0.004 (0.011)	0.003 (0.002)
N	216417	213634	193775	297753

*Note: Estimated effects are to show the difference between a sibling who had access to 4G-Internet and a sibling who did not. Standard errors are clustered at municipality level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, Norwegian born dummy, and high-income parents dummy, and control for municipality fixed effects. *p<0.10, **p<0.05, ***p<0.01.*

The estimates in Table 11 are all positive, but none of them are statistically significant. We can compare them to the estimates we found earlier when using coverage as a continuous variable. In column 1, GPA in middle school, the estimate is nearly identical to the estimate from Table 2, which was 0.016 with a standard error of 0.015. The estimate in column 2 differs a bit from the estimate from Table 4, being up from 0.004, with a standard error of 0.016. This can be caused by the dummy creating a “jump”, rather than the gradual change when coverage was continuous. The estimate for high school completion in column 3 has also changed. Unlike high school GPA the estimate has gone down, from 0.019 with a standard error of 0.010. As this regression now consists of only dummy variables, it could lead to difficulty finding covariation, hence why the estimate is now so low. While we do not have a previous estimate to compare column 4 with, we can see the estimate is slightly positive, but not statistically significant.

6.7 Possible confounder: Role of schools

Until now, we have controlled for municipality fixed effects, as it is highly likely that which municipality an individual lives in matters for their outcomes. However, we can narrow it down a bit further. As there are effects based on municipality, there could also be effects based on which school the individual went to within the municipality. We can run a regression to look at this, now grouping the individuals based on which middle school they went to, rather than which municipality they grew up in.

Table 12: Middle School fixed effects

	(1) MS GPA	(2) HS GPA	(3) HS-completion by age 19	(4) Any child before age 20
Cov	0.017	0.002	0.014	0.006^{***}
	(0.016)	(0.019)	(0.010)	(0.001)
N	215949	210748	189726	217668

*Note: Estimated effects are to show the difference between a sibling who had access to 4G-Internet and a sibling who did not. Standard errors are clustered at middle school level. Regression is based on 411 municipalities over 18 years. All regressions include year dummies, Norwegian born dummy, and high-income parents dummy, and control for municipality fixed effects. *p<0.10, **p<0.05, ***p<0.01.*

The estimates included in Table 12 are consistent with our previous findings, and the estimates for educational attainment are very close to those we found on municipality-level. The only noticeable difference is in column 4, the estimated effect on having a child before turning 20 is positive and clearly statistically significant. While this in itself is a strong result, our overall findings are too inconsistent to draw any conclusions. The estimate could be biased due to a low number of individuals becoming parents before turning 20.

6.8 Determinants of the LTE-network expansion

Finally, we can also take a look at how the expansion was decided, by looking at the characteristics of municipalities which were hit by the expansion in the same year. To do this, we will use average income, population, and a dummy for whether the municipality is rural. This will give us some numbers on how the expansion was decided.

Table 13: Characteristics of the expansion

	(1) First year of coverage
Average income	-0.0345
(per 100.000 NOK)	(0.0547)
Total population	-0.000739
(per 1000)	(0.000546)
1 [Rural]	0.713^{***}
	(0.056)
N	428

*Note: Estimates show change in the first year of LTE-coverage based on changes in municipality characteristics. Standard errors are robust. Regression is based on all 428 municipalities. *p<0.10, **p<0.05, ***p<0.01.*

The estimates included in Table 13 show the correlation between a few municipality characteristics and the first year the municipality received LTE-network coverage. It shows higher average income and population in the municipality negatively correlates with the year. This means richer, more populous municipalities received coverage earlier, but the estimates are not statistically significant. Additionally, the rural variable is equal to one if the municipality is rural, and we see it correlates with getting coverage a fair bit later than urban municipalities. This can also be shown graphically, shown in Figures A1-A3 in the Appendix.

7. Conclusion

In this paper, we have studied the influence of Internet access on the educational attainment of youth. The staggered expansion of the LTE-network in Norway has worked as an exogenous change to create a natural experiment. Using different empirical approaches and outcome variables has provided us with a broad view of the situation.

The initial regressions gave generally positive, but not statistically significant, estimates of the effect on grades in middle school and high school. It also estimated a statistically significant positive effect on high school completion, and a negative effect on having a child before the age of 20. Testing the two latter using a probability model supports this result.

Following this, using a mother fixed effect approach, the estimated effect was negative when comparing siblings when it comes to grades, while the effect on high school completion was not significant. However, based on the birth order effect, we expect the younger siblings to have lower educational attainment than their older siblings. Our results are not strong enough to disregard this effect, and therefore make it difficult to argue for a negative effect.

When allowing for gender-specific effects, we find that the grades of young men in both middle school and high school are negatively affected by Internet access. However, their high school completion rate is affected positively. These findings suggest that young men are negatively affected in the form of reduced time and effort put into schoolwork, but that this effect does not transfer to completion rate. Rather, it might help keep them in school by using the Internet as a social platform. It could also be caused by a negative effect on young women, suggested by some literature discussed in this paper.

This analysis provides some insight into the effects of the Internet, and social media's place in the everyday life of youth. Additionally, it contributes an empirical addition to a discussion mostly based on anecdotal evidence and correlative work. While the effect on educational attainment is still difficult to determine, the change in Internet access does affect youth through many different avenues and should not be left unchecked.

For further research, I would like to see work done using health data as the outcome variables, for both physical and mental health. Effects on educational attainment is only one side of the story, and diving into the health benefits or drawbacks could supplement the discussing greatly. It could also be interesting to use data on specific uses of Internet such as social media or online gaming, in order to separate the different effects of increased Internet access.

References

Aguiar, M., Bils, M., Charles, K. K., & Hurst, E. (2021). Leisure Luxuries and the Labor Supply of Young Men. *Journal of Political Economy*, *129*(2), 337-382. https://doi.org/10.1086/711916

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629-676. https://doi.org/https://doi.org/10.1257/aer.20190658

Allcott, H., Gentzkow, M., & Song, L. (2022). Digital addiction. *American Economic Review*, *112*(7), 2424-2463. https://doi.org/10.1257/aer.20210867

Alpert, W. T., Couch, K. A., & Harmon, O. R. (2016). A randomized assessment of online learning. *American Economic Review*, *106*(5), 378-382.

Angrist, J. D., & Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton university press.

Bhuller, M., Havnes, T., Leuven, E., & Mogstad, M. (2013). Broadband internet: An information superhighway to sex crime? *Review of Economic studies*, *80*(4), 1237-1266. https://doi.org/https://doi.org/10.1093/restud/rdt013

Bianchi, N., Lu, Y., & Song, H. (2022). The effect of computer-assisted learning on students' long-term development. *Journal of Development Economics*, *158*, 102919.

Bjørkeng, P. K. (2009). Ingen HD i bakkenettet.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, *120*(2), 669-700.

Braghieri, L., Levy, R. e., & Makarin, A. (2022). Social media and mental health. *American Economic Review*, *112*(11), 3660-3693. https://doi.org/doi.org/10.1257/aer.20211218

Bursztyn, L., & Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, *9*, 131-153.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, *3*, 1801-1863. https://doi.org/https://doi.org/10.1016/S1573-4463(99)03011-4

Carlin, B. I., Jiang, L., & Spiller, S. A. (2018). Millennial-Style Learning: Search Intensity, Decision Making, and Information Sharing. *Management Science*, *64*(7), 3313-3330. https://doi.org/10.1287/mnsc.2016.2689

Chen, Y., & Yang, D. Y. (2019). The Impact of Media Censorship

1984 or Brave New World? *The American Economic Review*, *109*(6), 2294-2332. https://www.jstor.org/stable/26737888

Chevalier, A., & Feinstein, L. (2006). Sheepskin or Prozac: The causal effect of education on mental health. *Available at SSRN 923530*. https://doi.org/http://dx.doi.org/10.2139/ssrn.923530

Cornaglia, F., Crivellaro, E., & McNally, S. (2015). Mental health and education decisions. *Labour Economics*, *33*, 1-12. https://doi.org/https://doi.org/10.1016/j.labeco.2015.01.005

Currie, J., Stabile, M., Manivong, P., & Roos, L. L. (2010). Child health and young adult outcomes. *Journal of Human resources*, *45*(3), 517-548. https://doi.org/10.3368/jhr.45.3.517

Egan, M., Daly, M., & Delaney, L. (2015). Childhood psychological distress and youth unemployment: Evidence from two British cohort studies. *Social Science & Medicine*, *124*, 11-17. https://doi.org/https://doi.org/10.1016/j.socscimed.2014.11.023

Eilertsen, T. (2018). Skjermbruk blant unge fortsetter å øke. https://www.aftenposten.no/norge/i/6nAvvr/skjermbruk-blant-unge-fortsetter-aa-oeke

Eines, T., Rosvoll, Å., & SørvikBrattøy, S. (2019). Hyppig bruk av sosiale medier kan gi ungdom psykiske utfordringer. *Sykepleien*. https://doi.org/10.4220/Sykepleiens.2019.79774

Figlio, D., Rush, M., & Yin, L. (2013). Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, *31*(4), 763-784.

Fletcher, J. M. (2008). Adolescent depression: diagnosis, treatment, and educational attainment. *Health Economics*, *17*(11), 1215-1235. https://doi.org/https://doi.org/10.1002/hec.1319

Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2011). The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review*, *101*(7), 2980-3018. https://doi.org/10.1257/aer.101.7.2980

Gerhardsen, M. H., Ole Petter. (2023). Barn blir klossete og syke av mye skjermbruk, Toppe! https://www.bt.no/btmeninger/debatt/i/O8WMe1/barn-blir-klossete-og-syke-av-mye-skjermbruk-toppe

Golin, M. (2022). The effect of broadband Internet on the gender gap in mental health: Evidence from Germany. *Health Economics*, *31*(S2), 6-21. https://doi.org/https://doi.org/10.1002/hec.4570

Hartman, E., & Hidalgo, F. D. (2018). An Equivalence Approach to Balance and Placebo Tests. *American Journal of Political Science*, *62*(4), 1000-1013. http://www.jstor.org/stable/26598797

Hvide, H. K., Meling, T. G., Mogstad, M., & Vestad, O. L. (2022). *Broadband internet and the stock market investments of individual investors*. http://www.nber.org/papers/w30383

Knapstad, M., Heradstveit, Ove & Sivertsen, Børge. (2018). *Studentenes helse- og trivselsundersøkelse 2018*. SHOT. https://siomediaservicesdatadev.blob.core.windows.net/studenthelse/SHoT_2018_Rapport.pdf?sp=r&st=2021-09-06T14:31:19Z&se=2099-09-06T22:31:19Z&spr=https&sv=2020-08-04&sr=b&sig=LrGPP8qXeLRHZGJBq0istgH4NbJMyxNQ3lb0Kg5W0rw%3D

Kommunikasjonsmyndighet), N. N. (2022). *Bredbåndsdekning 2022*. NKOM. Retrieved 10.05.23 from

Kotsbakk, K. K. N. T., Jøte. (2019). Skjermbruk kan hemme barns hjerneutvikling. https://www.nrk.no/viten/ny-forskning-fra-ntnu-og-usa-viser-at-skjermbruk-kan-hemme-barns-kognitive-og-sosiale-evner-1.14768801

Krokstad, S., Weiss, D. A., Krokstad, M. A., Rangul, V., Kvaløy, K., Ingul, J. M., Bjerkeset, O., Twenge, J., & Sund, E. R. (2022). Divergent decennial trends in mental health according to age reveal poorer mental health for young people: repeated cross-sectional population-based surveys from the HUNT Study, Norway. *BMJ open*, *12*(5), e057654.

Levine, D. I., & Painter, G. (2003). The Schooling Costs of Teenage Out-of-Wedlock Childbearing: Analysis with a Within-School Propensity-Score-Matching Estimator. *The Review of Economics and Statistics*, *85*(4), 884-900. https://doi.org/10.1162/003465303772815790

Lunde, M. (2023). Vi må gå sammen om å få ned skjermbruken. https://www.vg.no/nyheter/meninger/i/0QoQaA/vi-maa-gaa-sammen-om-aa-faa-ned-skjermbruken

Medietilsynet. (2008). *Barn og medier 2008: En karlegging av 8 til 18-åringers bruk av digitale medier*. Medietilsynet. https://www.medietilsynet.no/globalassets/publikasjoner/barn-og-medier-undersokelser/2008-barn-og-medier-8---18-ar.pdf

Medietilsynet. (2018). *Stadig flere barn får mobiltelefon*. Medietilsynet. Retrieved 10.05.23 from

Medietilsynet. (2020). *Barn og medier 2020: 9-18-åringers medievaner*. Medietilsynet. https://www.medietilsynet.no/globalassets/publikasjoner/barn-og-medier-undersokelser/2020/201015-barn-og-medier-2020-hovedrapport-med-engelsk-summary.pdf

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, *109*(4), 1426-1460.

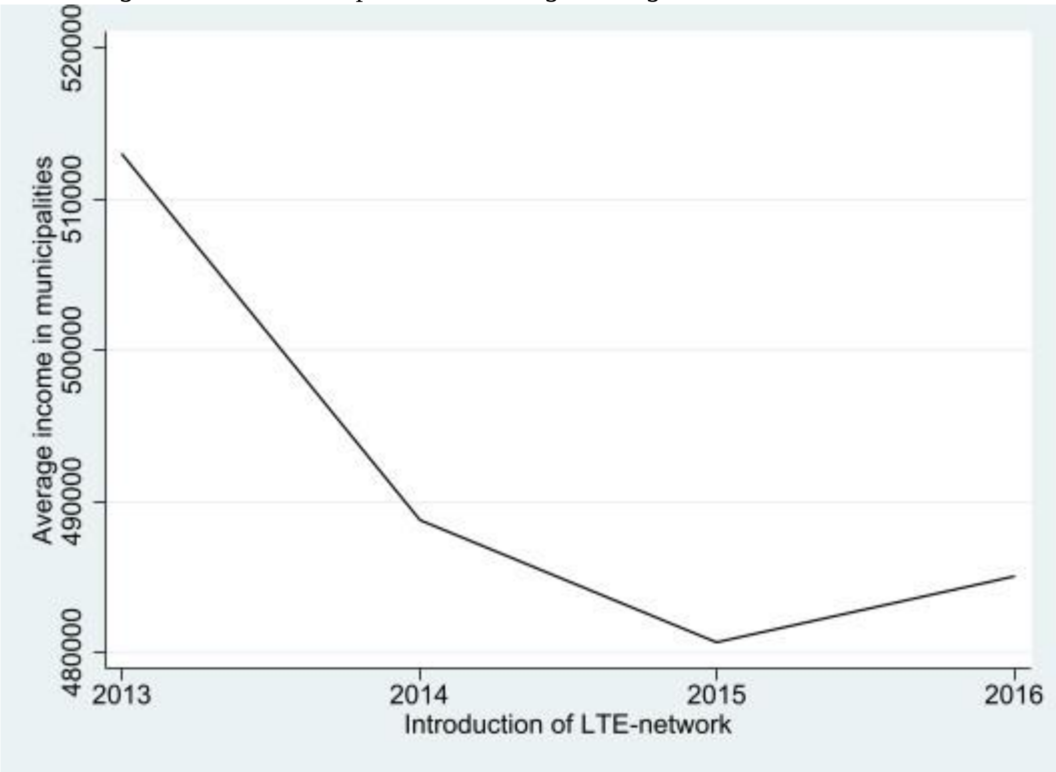
Phychiatry), A. A. A. o. C. A. (2017). *Teen Brain: Behavior, Problem Solving, and Decision Making*. Retrieved 10.05.23 from https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/The-Teen-Brain-Behavior-Problem-Solving-and-Decision-Making-095.aspx

Serra-Garcia, M., & Gneezy, U. (2021). Mistakes, Overconfidence, and the Effect of Sharing on Detecting Lies. *American Economic Review*, *111*(10), 3160-3183. https://doi.org/10.1257/aer.20191295

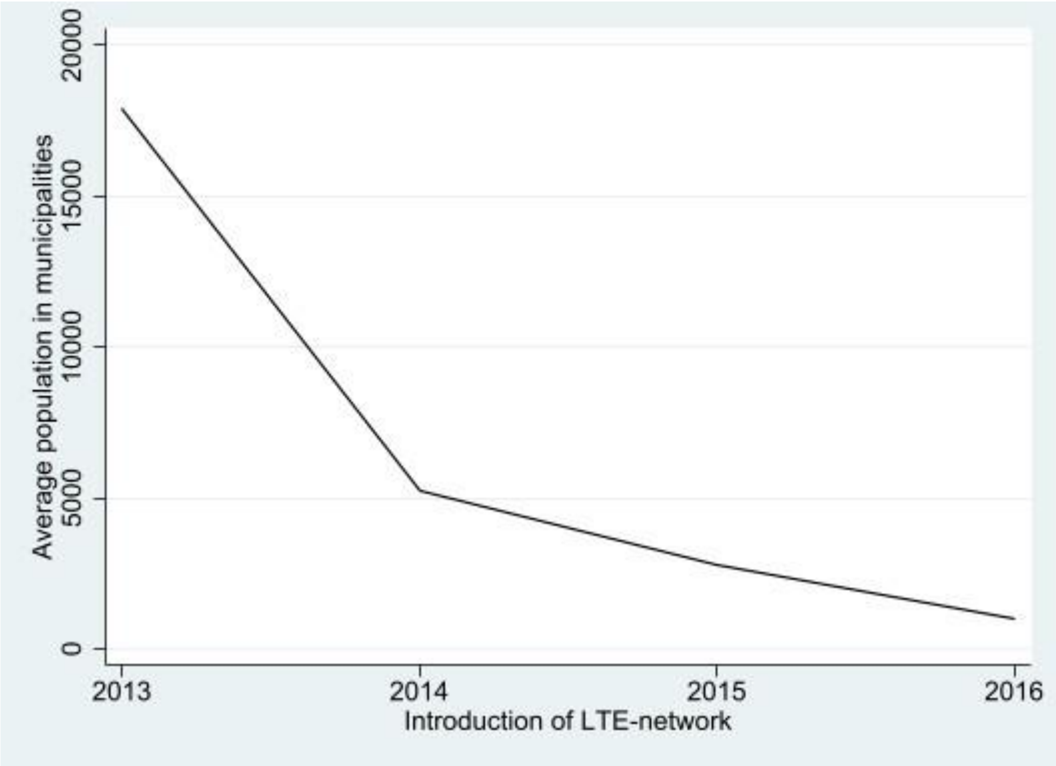
Sivertsen, B. S. J., Malin. (2022). *Studentenes helse- og trivselsundersøkelse 2022*. SHOT. https://studenthelse.no/SHoT_2022_Rapport.pdf
Smith, J. P., & Smith, G. C. (2010). Long-term economic costs of psychological problems during childhood. *Social Science & Medicine*, *71*(1), 110-115. https://doi.org/https://doi.org/10.1016/j.socscimed.2010.02.046
Surén, P. (2018). Har ungdommer dårligere psykisk helse enn før? *Tidsskrift for Den norske legeforening*.
Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents after 2010 and links to increased new media screen time. *Clinical Psychological Science*, *6*(1), 3-17.
Valkenburg, P. M., Adrian; Beyens, Ine. (2022). Social media use and its impact on adolescent mental health: An umbrella review of the evidence. *Current opinion in psychology*, *44*(1), 58-68. https://doi.org/https://doi.org/10.1016/j.copsyc.2021.08.017
van Rooij, A. J., Schoenmakers, T. M., van de Eijnden, R. J. J. M., & van de Mheen, D. (2010). Compulsive Internet Use: The Role of Online Gaming and Other Internet Applications. *Journal of Adolescent Health*, *47*(1), 51-57. https://doi.org/https://doi.org/10.1016/j.jadohealth.2009.12.021
Velsand, M. (2020). Barn og skjermtid: Skjermens vederstyggeligheter. https://www.vg.no/nyheter/meninger/i/mB6Jgv/barn-og-skjermtid-skjermens-vederstyggeligheter

Appendix

A1: Average income in municipalities receiving coverage



A2: Average population in municipalities receiving coverage



A3: Share of rural municipalities receiving coverage