The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0143-7720.htm

The effect of social network sites usage on absenteeism and labor outcomes: longitudinal evidence

Jie Chen
School of Finance and Economics, Jiangsu University, Zhenjiang, China

Received 9 June 2021
Revised 27 October 2021
28 November 2021
15 December 2021
19 December 2021
28 December 2021
Accepted 28 December 2021

Abstract

Purpose – The purpose of this paper is multifold. First, it is to investigate the relationship between social network sites (SNSs) usage and youth's school absenteeism. Second, it is to identify causal relationship between SNSs usage and absenteeism. Third, it is to explore whether SNSs usage causally affects youth's study–work choice after leaving high school. In addition to SNSs usage in general, abnormal SNSs usage is further discussed.

Design/methodology/approach – The Longitudinal Study of Australian Children (LSAC) data are utilised. Lagged variable analysis is used to alleviate reverse causality. Instrumental variable approach and the Lewbel method are used to identify causality. Random effects panel data approach (without and with IVs) is additionally applied to increase efficiency and account for individual-specific effects. Random effects approach allowing for within and between effects is applied, enabling us to control for fixed effects. The primary instrument is a dummy indicating whether a youth more often communicates with close friend electronically or face-to-face.

Findings – Using SNSs leads to significantly higher probability of a teenager being late for school, skipping class and having trouble not following school rules. The effect is more consistent regarding abnormal SNSs usage, compared to SNSs usage in general. Additionally, SNSs usage decreases the probability of a youth studying after 18 years old, even after controlling for absenteeism.

Practical implications – The findings in this paper highlight the importance of preventing youth (e.g. via enabling children-safe mode or setting up maximum daily access time) from overusing SNSs.

Social implications – With the transition to hybrid (mixing remote and face-to-face) learning during and after COVID-19, online interactions are becoming inevitable in students' learning. The findings in this paper indicate that usage, especially abnormal usage, of SNSs increases the probability of absenteeism call for attention from stakeholders including teachers, parents and youth themselves.

Originality/value – This paper provides the first causal and longitudinal evidence linking SNSs usage to absenteeism and youth labor outcomes.

Keywords Youth, Social network sites (SNSs), Social media, Absenteeism, Labor outcomes, Longitudinal, Causality

Paper type Research paper

1. Introduction

The usage of social network sites (SNSs) is pervasive around the world, especially among youth. According to Statistica, the three most popular SNSs worldwide as of January 2021 are Facebook, YouTube and WhatsApp [1], with Facebook itself has 2.7 billion active users. Statistica also finds that, in Australia, 94% of the 12–24 user group are on SNSs. The Pew

JEL Classification — J13, J21, J24

This study uses data from Growing up in Australia: the Longitudinal Study of Australian Children (LSAC). The LSAC is conducted in partnership between the Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS), with advice provided by a consortium of leading researchers from research institutions and universities throughout Australia. This paper is part of the Global Labor Organization (GLO) Virtual Young Scholars Program.

Funding: Project name – Research on the Structural Reformal Motivation and Growth Potential of Jiangsu Cultural Industry under the New Normal. Funding agency – Key Projects of Philosophy and Social Sciences in Jiangsu Universities (2017ZDIXM037).

International Journal of Manpower

0143-7720
DOI 10.1108/IJM-06-2021-0338

Research Center finds that around 76% of teens aged 13–17 are on social media (Lenhart *et al.*, 2015). Yet the effect of SNSs usage on youth school performance and labor outcomes is understudied. This is surprising since Australia youth labor market belongs to the most studied Anglo-Saxon youth labor market system (Pastore, 2015, 2018). It has been documented that the school-to-work transition experience of youth is dramatically heterogeneous due to uncontrollable reasons yet can greatly impact their later career trajectories (Pastore and Zimmermann, 2019; Pastore *et al.*, 2021). Digital competency is an important skill for the current labor market. Understanding the relationship between SNSs usage and youth school performance and labor outcomes can shed light on channels that potentially influence the school-to-work transition of youth.

SNSs are online platforms through which individuals construct public (or semi-public) profile and connect with other people who share similarities (e.g. interests, backgrounds) or are related in real life (Boyd and Ellison, 2007). SNSs are also referred to as social networking sites. In recent years, SNSs have been frequently referred to as or used interchangeably with "social media" (Obar and Wildman, 2015; Swist *et al.*, 2015). Due to the fast-evolving technology development, it is almost impossible to track the precise origin and meaning of these terminologies. In this paper, we do not differentiate between these terminologies and stick to SNSs in the exposition.

Many controversies surround children and teenagers' use of SNSs. On the one hand, SNSs provide many unique benefits to youth. For example, some believe it is necessary to equip youth with technical skills at an early age, so that youth can become technically adept to navigate freely on the Internet and learn to be a good "netizen". SNSs also give youth a chance to socialize with their friends remotely, especially when a child does not have many companions nearby (e.g. during COVID period). SNSs can provide a sense of community and belonging to isolated or vulnerable youth (Swist *et al.*, 2015). On the other hand, SNSs can expose youth to invisible challenges such as online abuse and child pornography. With the quick transition and rapid adoption of online learning (especially after COVID), during which youth inevitably connect with peers using SNSs, it is of urgent policy interest to untangle the effect of SNSs usage on youth outcomes. Yet, causal evidence on the related topic is scarce, even less discusses the mechanism, if at all (Keles *et al.*, 2020).

Prior to the transition to remote learning due to COVID-19, limiting children's screen time appeared a legitimate decision by parents. However, as education institutions worldwide are now forced to offer remote teaching, which shows no sign of reverting back to the traditional classroom teaching in the years to come, understanding whether and to what extent children's SNSs usage affects their academic and labor outcomes can at least help families and the government to start preventative actions at an early stage if necessary.

In this paper, we use a longitudinal dataset from Australia to analyze the effect of SNSs usage on teenagers' school absenteeism, as well as their study–work choices upon adulting. We find that using SNSs leads to higher likelihood of a student being late from school, skipping classes, being absent and having trouble not following school rules. Meanwhile, teenagers who use more SNSs in early years are more likely to be out of school after 18 years of age. Note that although this study shows that SNSs usage relates to higher probability of youth not studying or not working, more research are needed before we can conclude whether these changes are a good or bad signal of youth's academic/job performance.

This paper has several major contributions. First, to the best of our knowledge, this paper is the first one that provides causal evidence regarding the impact of SNSs usage on young people's education and labor outcomes. Second, we control an extensive set of covariates, including youth's academic performance in previous years, income group of the family, youth's drug usage history etc. Third, we introduce two novel instruments. The primary instrument is whether a teenager more often communicates with close friends electronically or face-to-face. The second one is the time spent on SNSs in a typical week in Wave 7 due to exogenous shocks. Fourth, we use a heteroskedasticity-based Lewbel method (Lewbel, 2012) to check the robustness of the IV results. Lastly, we apply random effects panel data approach on the longitudinal data, which allows individual-specific effects to be controlled, to analyze the effect of SNSs usage on youth outcomes.

The remainder of this paper is structured as follows. Section 2 provides the literature review. Section 3 provides the research context, introducing the dataset and the research questions. Section 4 discusses the variables of interest and the methodology. Results are discussed in Section 5. Section 6 is for the conclusion.

2. Literature review

Early investigation of SNSs usage frequently focused on Internet usage in general, but not SNSs usage in specific. For example, Kraut *et al.* (1998) conduct a two-wave longitudinal survey among 93 Pittsburgh families and find that Internet usage negatively relates to social involvement or psychological well-being. Although the SNSs software is included in their investigation, the authors do not isolate the pure effect of SNSs usage. Interestingly, a revisit of the same topic in Kraut *et al.* (2002) reveals the opposite that Internet usage positively relates to social involvement and well-being. The contrasting findings in these two studies highlight the importance of providing updated research evidence. More recently, Brandtzæg (2012) studies the relationship between SNSs usage and social capital accumulation. The author utilizes a three-wave dataset on Norwegian online users aged between 15 and 75 to study the effect of SNSs usage on social capital accumulation. The author finds that SNSs users score higher in three out of four social capital dimensions, including face-to-face interactions with close friends, offline acquaintances and connections with socially heterogeneous groups.

Yet, SNSs users are found to report feeling lonely more frequently than non-SNSs users.

In contrast to the lack of studies directly related to SNSs usage, SNSs have penetrated exponentially into every aspects of our life since their introduction in the 1990s. Notably, SNSs usage, being a subset of Internet usage in general, has unique features that distinguish itself from other Internet usage activities such as browsing through news or searching for information. These unique features include but are not limited to social stickiness, interpersonal connectivity, social isolation and comparison etc. (Ostic *et al.*, 2021; Durak, 2020). According to Durak (2020), although some researchers tend to view problematic social media usage as a subset of problematic Internet usage, the author finds no significant correlation between problematic Internet usage and problematic social media usage. Hence, it is important to view problematic social media usage research as a separate type from problematic Internet usage research. More generally, problematic Internet usage includes but is not limited to computer addictions, online gambling addictions, compulsive online shopping and problematic social media usage (Durak, 2020; Young and Case, 2004). With the popularization of SNSs, an increasing body of literature investigates the relationship between SNSs usage and other outcomes, such as school performance or labor outcomes. On the one hand, SNSs usage, or digital device usage in general, is linked to impaired work performance, deteriorated school performance, impaired physical health condition and mental health problems such as anxiety and depression (Baert *et al.*, 2020; Amez *et al.*, 2019; Andreassen, 2015; Keles *et al.*, 2020; Posso, 2016; Kreski *et al.*, 2021). On the other hand, SNSs usage has been linked to quick spread of knowledge, free learning opportunities, a sense of belonging and community, higher social capital accumulation and rich recreational activities (Allam *et al.*, 2012; Lim, 2013; Swist *et al.*, 2015; Brandtzæg, 2012). Thus, SNSs usage is a double-edged sword which needs to be better understood.

When it comes to absenteeism, no previous study has explored the effect of SNSs usage on absenteeism. Topic-wise, the most closely related paper is by Austin and Totaro (2011).

The authors measure absenteeism as the number of days, in the 30 days preceding their survey, that a student is absent from school due to illness (or injury) or skipping. The independent variable of interest in Austin and Totaro (2011) is Internet usage, which is different from our independent variable of interest SNSs usage. The authors categorize the Internet usage variable based on the place in which the Internet is accessed: the usage is *intense* if the Internet is accessed at multiple places including home, cafe and other places; the usage is *moderate* if the Internet is only accessed at home; and the usage is *light* if the Internet is only accessed at a library. As can be seen, the treatment variable of interest in Austin and Totaro (2011) is Internet usage in general, whereas our study is focused on the effect of SNS usage specifically. Notably, back when Austin and Totaro (2011) published their paper, SNSs were not as popular and many of the currently popular SNS platforms (e.g. Snapchat, Tiktok) were yet to be released.

One recent paper exploring similar research questions is by Amez *et al.* (2019). The authors conducted a longitudinal study to explore the effect of smartphone usage on university students' academic performance. They find that increased smartphone usage causes a significant decrease in course marks and a significantly lower probability of passing exams. Yet, the authors note that their paper suffers one major limitation: they are not able to untangle the mechanisms behind the negative relationships. Meanwhile, it is apparent that their study is focused on smartphone usage, but not SNSs usage *per se*.

Despite the aforementioned studies, there is a lack of studies focusing specifically on SNSs usage, less evidence is longitudinal or causal in this literature (Baert *et al.*, 2020; Keles *et al.*, 2020). In the systematic review done by Keles *et al.* (2020), the authors find that 12 out of 13 reviewed studies used cross-sectional data and identify the challenges faced by cross-sectional studies. The authors point out the importance of identifying the mechanisms behind the putative effects and call for more longitudinal studies. The authors also highlight the lack of causal evidence in this area. Amez and Baert (2020) recently reviewed the literature related to smartphone usage and academic performance. The authors identify the lack of causal evidence as a major limitation in this area. They thus suggest collecting longitudinal data to better control for individual fixed effects in regression analyses. Our paper echoes the authors' appeal in multiple facets. We utilize multiple identification strategies to provide causal evidence on the effect of SNSs usage on youth outcomes. We also identify that, after controlling for absenteeism, SNSs usage affects youth's later school–work choices.

The scarcity of causal evidence related to SNSs usage, contrasting the upsurge of study and work-digitization, necessitates new studies on related topics. Especially after the transition to remote learning due to COVID-19, understanding and quantifying the impact of SNSs usage on youth education and labor outcomes is of urgent policy interest.

3. Research questions and the dataset

3.1 Research questions

The longitudinal nature of the dataset allows us to avoid the reverse causality problem on the one hand and track the effect of SNSs usage on teenagers' education and labor outcomes over time on the other hand.

The first main research question is whether SNSs usage affects teenagers' absenteeism at school. To answer this question, the SNSs usage variable is taken as the treatment variable. Then, the dependent variables are five absenteeism measures (details are provided in Section 4.1). The second main research question is whether SNSs usage at teenage affects young adults' study–work choices. The treatment variable remains unchanged. The dependent variable is now a dummy indicating whether a subject is studying or not or at the time of the Wave 8 survey, when teenagers are 18–19 years old. Understanding the study–work status of high-school graduates matters because school-to-work transition is a topic of heated policy interest (Pastore, 2015). How smooth the transition is largely affects the career trajectory of individuals. Different transition patterns have different policy implications. For example, if we find that an overwhelming proportion of individuals are in the status of neither working nor studying, further action is needed to investigate why this is the case and whether any additional supportive training is needed from the government.

Along with these two main research questions, we also want to explore whether absenteeism mediates the effect of SNSs usage on latter school–work choices. After all, even if SNSs usage is found to significantly affect school–work choices, it is possible that the effect is channeled partially through changing students' school attendance. For example, a student who uses SNSs more often may be absent from school more often, which consequently increases the likelihood of the student not studying at 18–19 years old.

3.2 The dataset

In this paper, we use data from the Longitudinal Study of Australian Children (LSAC). The LSAC is a biannual study that commenced in 2004. It follows two parallel and unrelated cohorts of Australian children: the B cohort (i.e. infant cohort) with children born in 2003–2004 and the K cohort (i.e. child cohort) with children born in 1999–2000. Both cohorts are selected based on a two-stage (i.e. postcode and children) clustered sample design, so that they are representative of all Australian children in their respective cohort. As of 2020, eight waves of data have been released, reporting information of the B cohort up to the age of 14–15 and information of the K cohort up to the age of 18–19. Only the data from the K cohort is used in this paper because study–work choice information is only available in this cohort as of 2020. Specifically, each wave of the LSAC survey is composed of a series of survey modules. Depending on the age and the stage of child development, modules are removed or added accordingly in different waves. Questions about the employment history of children belong to a module called Event History Calendar (EHC). The EHC module captures children's employment, study and residential history and is added to the LSAC survey for the K cohort starting in Wave 7.

In Australia, primary and secondary education is compulsory for children aged between six and sixteen [2]. According to Youth Law Australia [3], full-time working is restricted in some Australian states for children under the age of 18. Because of this, we exclusively focus on the employment status of the K cohort subjects, who turned 18 before the Wave 8 interview. Specifically, the most recent three consecutive waves of the K cohort of the LSAC data collected between 2014 and 2018 are utilized in this paper. Within the K cohort, unmatched observations across waves are dropped, leaving a total of 2,673 observations.

Extensive information about the sampled teenagers is available in the linked datasets, ranging from basic demographics to school performance to general behavioral proxies. Importantly, in Wave 6 and 7, respondents were asked the following question: "How often do you use a computer or computer-like device to spend time on social networking sites?" The computer-like devices refer to any of the following: Tablet, Computer, Laptop, Smartphone, Keypad phone, Portable Media Device (e.g. iPod), Gaming console, Handheld gaming device, TV, Smart TV. The available answers are "Almost every day/Once or twice a week/A few times a month/Once a month or less/Never". The variable forms the primary treatment variable of interest in this paper. Normally, categorical independent variables are collapsed into dummies to facilitate interpretation. However, given the scarce literature exploring the effect of SNSs usage [4], it is unclear which way is more appropriate to collapse these five categories, or whether these five categories should be collapsed at all. If we are to follow a strong hypothesis that any kind of SNSs usage hurts students' school attendance, then it is more appropriate to set the "Never" category to zero and the remaining four categories as one. Yet, if we follow a weak hypothesis that only abnormal usage of SNSs is harmful to attendance, then only “Almost every day” should be set to one and the remaining four categories should equal zero. In light of this issue, we explore both specifications in Section 5.

When it comes to the decision to continue studying, one important determinant is students’ academic performance. However, reliable and comparable academic performance measures are often not available in most studies. This is not the case with the LSAC data. The LSAC data are linked to NAPLAN data, allowing me to control for students’ academic performance over the years. NAPLAN, which stands for the National Assessment Program -Literacy and Numeracy, is a nationwide assessment for Australian students in Years 3, 5, 7 and 9. The program was introduced in 2008 and happens annually. It evaluates four separate learning areas: reading, writing, language conventions (i.e. spelling, grammar and punctuation) and numeracy. The purpose of NAPLAN is to help stakeholders (e.g. schools, parents and students) understand how students are progressing relative to the national average of their cohort, so that stakeholders can tailor their later effort accordingly. The assessment is formative, not summative, in that it merely measures performance but does not pass or fail any student. The NAPLAN score roughly ranges from zero to 1,000.

4. Methodology

4.1 Dependent variables

As mentioned before, the first set of dependent variables of interest are teenage absenteeism measures. A total of five absenteeism measures are available in the dataset: in the 6 months preceding the survey, how often a teenager was late for school/skipped classes/was absent from school with parent permission/was absent from school without parent permission/got into trouble for not following school rules. These five variables are largely in line with the absenteeism measures used by Austin and Totaro (2011). All the absenteeism measures are collapsed into dummy variables to account for potential recall bias. Recall bias is a commonly criticized source of measurement error in survey data. Generally, there is a tradeoff between the response accuracy and the length of period over which subjects need to recall (Clarke *et al.*, 2008). Although the optimal length of period exists theoretically, it is susceptible to all kinds of influence factors in practice. To minimize related recall bias, we collapse the categorical variables into dummies which equal zero for “never” and one for all positive frequencies. The logic is that respondents may not accurately recall the days corresponding to a specific type of absenteeism after a while, but they should be able to recall whether a specific type of absenteeism happened or not. Note that some of the absenteeism measures highlight a choice (e.g. skipping classes), whereas some highlight a circumstance leading to an outcome (e.g. having trouble not following school rules). The difference between the different types of absenteeism measures has implications on the results interpretation. Specifically, “being absent with parent permission” can be a good placebo test of the other absenteeism measures. Since it is extremely unlikely that parents allow children to not go to school just because children have used too much SNSs, SNSs usage is expected to have no robust effect on this absenteeism measure. This is indeed what we observe in Table 1. Meanwhile, “being late for school” and “having trouble not following school rules” are likely circumstances resulting from SNSs usage. For example, a teenager may use SNSs for too long at night to sleep well. Deteriorated sleep in turn results in late wakeup (consequently being late for school), poor memorizing and emotional disturbances (consequently high likelihood of getting into school trouble). We will delve deeper into the sleep-related variables in Section 5.4.

As can be seen from Table 2, five answers were available for each question, ranging from “never” to “10 or more times”. For the ease of interpretation, these categorical absenteeism measures are collapsed into binary variables in the econometric analysis. We define the “never” category as 0 and the other four categories as 1. That means whenever a teenager has non-zero frequencies recorded, he is regarded as a truant in the analysis. Although

Dep. var.	Late for school (1)	Skip classes (2)	Absent with permission (3)	Absent without permission (4)	Trouble not following rules (5)
Panel A: results with weak hypothesis (only abnormal usage is harmful)					
Baseline	0.079***	0.078***	0.061***	0.027**	0.077***
	(0.020)	(0.018)	(0.017)	(0.014)	(0.021)
N	2,223	2,223	2,223	2,222	2,225
Full	0.041*	0.058***	0.039**	0.011	0.066***
	(0.023)	(0.020)	(0.019)	(0.015)	(0.024)
N[a]	1,791	1,791	1,791	1,790	1,793
IV	0.892***	0.543**	0.331*	0.389**	0.700**
	(0.317)	(0.245)	(0.196)	(0.184)	(0.291)
IV F-statistic	16.35	16.35	16.68	16.50	16.72
Durbin p-value	0.000	0.023	0.114	0.016	0.010
N[a]	1,772	1,772	1,772	1,771	1,774
Lewbel (χ^2 of BP	0.212**	0.140*	0.048	0.016	0.074
test: 35.28)	(0.088)	(0.079)	(0.070)	(0.055)	(0.091)
N[a]	1,772	1,772	1,772	1,771	1,774
Panel B: results with strong hypothesis (any kind of usage is harmful)					
Baseline	0.093***	0.122***	0.012	0.063***	0.135***
	(0.031)	(0.022)	(0.024)	(0.016)	(0.029)
N	2,223	2,223	2,223	2,222	2,225
Full	0.091**	0.080***	0.008	0.033*	0.126***
	(0.035)	(0.026)	(0.028)	(0.019)	(0.033)
N	1,786	1,786	1,786	1,785	1,788
IV	1.582**	0.964**	0.597	0.696**	1.253**
	(0.616)	(0.465)	(0.372)	(0.348)	(0.562)
IV F-statistic	10.40	10.40	10.26	10.31	10.45
Durbin p-value	0.000	0.021	0.077	0.019	0.010
N	1,772	1,772	1,772	1,771	1,774
Lewbel (χ^2 of BP	0.137***	0.069*	0.066	0.030	0.107**
test: 216.68)	(0.053)	(0.041)	(0.046)	(0.026)	(0.051)
N	1,772	1,772	1,772	1,771	1,774

Note(s): All the dependent variables are dummies which equal 0 for "never" and equal 1 for positive frequencies. Baseline regressions only include SNSs usage as the control variable. Full regressions additionally control for other demographics as those appear in Table 7. IV stands for the instrumental variable approach estimates. Lewbel stands for the heteroskedasticiy-based IV estimates. Robust standard errors are reported in parentheses.

Table 1.
Absenteeism results

absenteeism is generally related with poor academic performance, it is possible that students are absent a few times due to unavoidable reasons (e.g. sickness, heavy traffic) (Austin and Totaro, 2011). Thus, classifying all positive frequencies as truancy is extra stringent. If anything, our specification of the absenteeism measures leads to under-estimation of the related coefficients. The second set of dependent variables is related to teenagers' study status. Typically, high-school graduates have three options: to continue studying; to start to work; or to remain idle. For those who decide not to pursue further education, they either work or remain idle. Depending on the supply and demand in the labor market, as well as individual competitiveness, the duration of school-to-work transition differs greatly among individuals. The duration matters, especially for fresh graduates. Research has shown *duration dependence* for those who remain idle after graduation. Duration dependence can be positive or negative, in the sense that those who remain unemployed for longer period of time become less or more competitive over time as their employability characteristics deteriorate or

Categories	Late for school	Skipped classes	Absent with parent permission	Absent without parent permission	Trouble not following school rules
Never	696	1,814	366	2,083	1,452
1–2 times	820	345	910	165	595
3–6 times	438	107	687	47	183
7–9 times	157	26	205	19	62
10 or more times	245	64	188	41	66
Total	2,356	2,356	2,356	2,355	2,358

Note(s): All five absenteeism measures are from Wave 7 of the LSAC data – when the subjects are 16–17 years old. Each column corresponds to one category of the five absenteeism measures

Table 2. Description of absenteeism in Wave 7

improve (Pastore *et al.*, 2020). Some might argue that not all young people are unemployed unwillingly and some youth intentionally choose to have a gap period. Regarding the impact of having a gap period, research evidence is also mixed. Some find that students with gap-period experience are more likely to drop out of a university degree (Parker *et al.*, 2015). Some find that students with gap-period experience demonstrate higher motivation level at university (Rose Birch and Miller, 2007). Echoing the existing literature on school-to-work transition, we explore the study–work choice patterns of fresh graduates post high school.

After 18 years old, it becomes legal for Australian children to stop going to school [5]. As mentioned before, literature has documented that the usage of SNSs can increase mental health problems facing young people, which further increases the likelihood of a student quitting school (Keles *et al.*, 2020). We thus wonder whether the decision of continuing education relates to SNSs usage. Since students are allowed to work part-time in most Australian states, we further classify the study/no study status depending on whether a teenager also works. That renders four categories: a teenager is studying but not working, a teenager is studying and working, a teenager is not studying but working, and a teenager is neither studying nor working at the time of the Wave 8 survey. The dependent variables used in the regressions are described below in Table 3.

Due to the categorical nature of the study–work choice variables, we use multinomial logistic (MNL) model instead of Ordinary Least Square (OLS) estimations. The MNL model is a semi-elasticity in the sense that log odds are regressed on linear combinations of the covariates. Additionally, the MNL model requires one category of the variable to be set as the baseline, so that the coefficients can be interpreted as the change in relative log odds (or

	Mean	SD	N
Panel A: absenteeism in wave 7			
Late for school (yes for nonzero outcomes)	0.705	0.456	2,356
Skip class (yes for nonzero outcomes)	0.230	0.421	2,356
Absent with parent permission (yes for nonzero outcomes)	0.845	0.362	2,356
Absent without parent permission (yes for nonzero outcomes)	0.115	0.320	2,355
School trouble (yes for nonzero outcomes)	0.384	0.487	2,358
Panel B: study-work choice in wave 8			
Study (yes = 0)	0.385	0.487	2,673
Study – no work	0.180	0.384	2,673
Study – work	0.435	0.496	2,673
No study – work	0.300	0.458	2,673
No study – no work	0.085	0.278	2,673

Table 3. Description of all dependent variables

relative risk ratio, depending on the situation) in response to a unit change in an independent variable.

4.2 Choice of control variables

In the baseline specification, we only include usage of SNSs as the control variable. The distribution of social network usage in Wave 6 and Wave 7 is presented below in Table 4. Table 4 shows that teenagers who use SNSs almost every day increased to more than 80% in Wave 7, as opposed to around 60% in Wave 6. The social media market is rather dynamic and evolves quickly. For example, Tiktok was not released until late 2016, yet kids aged 4–15 already spend 80 min daily on Tiktok by 2020, being comparable to the 85 min on Youtube, according to Qustodio [6] - a digital safety app maker. Note that the frequencies in the table reflect SNSs usage in specific but not mobile phone usage in general. For example, a teenager may use the mobile phone to play video games or do homework-related tasks every day, which generally will not be reflected in the frequency table. There is a possibility that teenagers game or watch videos via SNSs, but the current survey questions do not allow us to identify specific activities that are done via SNSs. In this sense, the SNSs usage variable encompasses all the activities that are done via SNSs [7]. SNSs usage from Wave 6 is used for two reasons. First, it allows us to look at outcome variables in two subsequent waves: absenteeism in Wave 7 and study–work choice in Wave 8. Second, SNSs usage becomes more common in Wave 7 (more than 90% of the sample than the 87.35% in Wave 6), making the corresponding estimations less precise.

To account for other potential confounders, we control for a comprehensive list of covariates in the full specification (See Table 5 for a complete list of the covariates). First, we control for basic demographic variables such as age and gender. We also control for the ethnicity background of the teenagers, including whether they are Australia's indigenous people. Third, the type of school that the subject attends is controlled. Australia has three major types of schools: government schools, independent schools and Catholic schools. According to ACARA (Australian Curriculum, Assessment and Reporting Authority), 70% of the total number of schools are government schools, 18.5% are Catholic schools and 11.6% are independent schools as of 2020 [8]. Government schools educate 65.6% of the students. Catholic schools and independent schools educate 19.4 and 15.0% of the students, respectively [9]. Additionally, the main parent's years of education is also controlled. One might notice that the sample sizes vary among different specifications, this is because of missing values in control variables and in the instrumental variables. To alleviate the concern of sample attrition bias, we carry out covariate-dependent missingness (CDM) test on the main covariates. We find that, although the missing values are not missing completely at random, the missing patterns are reasonably random given the auxiliary variables *age* and *gender*. Hence, the missing values will unlikely distort or bias our estimates.

Different education qualifications are converted to their corresponding years of education based on a mapping of the Australian education system onto the International Standard

	Wave 6		Wave 7	
Category	Frequency	Proportion	Frequency	Proportion
Almost every day	2,143	0.646	2,428	0.824
Once or twice a week	539	0.162	275	0.933
A few times a month	122	0.037	58	0.197
Once a month or less	114	0.034	51	0.173
Never	399	0.120	135	0.458
Total	3,317	1	2,947	1

Table 4. Frequency of SNSs usage

Variable	Definition
Age	Age in years
Female	Gender dummy: male = 0, female = 1
Indigenous	Indigenous status dummy: non-indigenous = 0, indigenous = 1
Country of birth	Categorical: Confidentialised, Australia, New Zealand, United Kingdom, India, United States of America, South Africa
Home language is English	Dummy: English = 0, not English = 1
Number of younger siblings	Number of younger siblings
Number of older siblings	Number of older siblings
Number of same-age siblings	Number of same-age siblings
Mother education	Mother's years of education
Peers aged <18 in linked area (%)	Percentage of peers aged below 18 in neighboring areas
IRSAD index	Index of Relative Socio-economic advantage, an ordinal socio-economic status measure
Parent away	Dummy:
Drug use	Have ever used drugs: yes = 1; no = 0
Alcohol	Have tried alcohol more than a few sips: yes = 1; no = 0
Boarding school	Attends boarding school: yes = 1, no = 0
Income group of the main parent	Categorical: less than $500 per week $25,999 or less per year, $500-$999 per week $26,000–$51,999 per year, $1,000-$1,999 per week $52,000–$103,999 per year, $2,000 or more per week $104,000 or more per year
State of residence	Categorical: NSW (New South Wales), VIC (Victoria), QLD (Queensland), SA (South Australia), WA (Western Australia), TAS (Tasmania), NT (Northern Territory), ACT (Australian Capital Territory)
School type	Categorical: Government school, Catholic school, independent school, not in school
NAPLAN scores – numeracy	Continuous from 0 to 1,000
NAPLAN scores – spelling	Continuous from 0 to 1,000
NAPLAN scores – writing	Continuous from 0 to 1,000
NAPLAN scores – grammar	Continuous from 0 to 1,000
NAPLAN scores – Reading	Continuous from 0 to 1,000

Note(s): NAPLAN stands for National Assessment Program -Literacy and Numeracy

Table 5.
Definition of covariates

Classification of Education [10]. The exact mapping is as follows: Certificate I or II is equivalent to 10 years of education; Certificate III is equivalent to 12 years of education; Certificate IV is equivalent to 14 years of education [11]; diploma or advanced diploma is equivalent to 16 years, so is bachelor's degree; Graduate Certificate or Graduate Diploma is equivalent to 18 years; and postgraduate degree is equivalent to 19 years of education.

To account for the influence of family background, we include the IRSAD – Index of Relative Socio-economic Advantage and Disadvantage – as a control variable. The IRSAD is one out of several Socio-Economic Indexes For Areas (SEIFA). SEIFA are a bundle of indexes prepared by the Australian Bureau of Statistics to measure the relative social-economic advantage and disadvantage of areas based on Census data. Specifically, IRSAD is a continuous rank of areas from the most disadvantaged to the most advantaged, where the social-economic advantage and disadvantage is "people's access to material and social resources and their ability to participate in society" (Australian Bureau of Statistics, 2011).

We have also tried including variables such as the country of birth of the teenager and behavioral proxies (e.g. attention span). However, none of these variables significantly, either economically or statistically, correlate with the outcome variables of interest. Further, including these variables decreases the explanatory power of the model. Thus, these variables are purposefully left uncontrolled in the current paper.

4.3 Instrumental variables

With the longitudinal dataset, the reverse causality problem is minimized. Yet, it is still possible that both SNSs usage and students' absenteeism are caused by other factors. To address potential endogeneity issues, existing research has explored various instrumental variables including one's phone contract, perceived quality of WIFI, average broadband speed [12] in each area (Amez *et al.*, 2019; McDool *et al.*, 2020). These variables are unfortunately not available in the current dataset. In this paper, we explore novel instruments and incorporate them into two different instrumental variable methods. One is the conventional IV method. The other is the Lewbel method – a heteroskedasticity-based IV method.

Two sets of instrumental variables are utilized. The main instrumental variable is whether a teenager more often communicates with her close friends face-to-face or electronically. The survey question reads, "Thinking about your close friends, how much of the time do you interact with them face-to-face or via electronic devices?" The available answers are as follows: All or almost all face-to-face; Mostly face-to-face; About half and half; Mostly via electronic devices; All or almost all via electronic devices. This categorical variable is converted into a dummy which equals zero for the first two categories and one for the other three categories. Intuitively, if a teenager tends to communicate with friends electronically, then her usage of SNSs must also increase correspondingly. Meanwhile, it is unlikely that the means of communication directly affect a student's absenteeism.

Additionally, the Wave 7 (i.e. when subjects aged 16–17) survey included two new questions: "How much of this weekday (weekend) online time is spent on social media? (This includes things like Facebook, Instagram, Twitter and Tumblr)". These two questions are coded into one variable: *timeSNS*. Specifically, *timeSNS* equals one if the respondent reported spending about half/more than half/all of the online time of the week, including weekday and weekend, on social media and zero otherwise. This variable is added as an additional instrument when evaluating the effect of SNSs usage on teenagers' outcomes at the age of 18–19.

The rational of choosing *timeSNS* as an additional instrument is as follows. In the period 2015–2016, three of the ten most popular SNSs in 2021 among Australian users entered the Australian market. These are Tumblr (entered in August 2015), Snapchat (entered in February 2016) and Tinder (entered in September 2016). As of December 2020, these three SNSs have a total of 14.1 million active users, close to the 16 million users on Facebook or Youtube, according to Civic Web Media [13]. Snapchat is even among the top 3 most popular social media among kids and teenagers, according to eSafety Commissioner of the Australian government [14]. Unambiguously, the entry of the three SNSs into Australia creates an exogenous shock to Australian youth's SNSs usage during the survey window (i.e. between Wave 6 and Wave 7). Youth will adjust their SNSs usage to accommodate the shock, which will be reflected in the time they spend on social media. Importantly, the adjustment varies depending on the extent to which a young person is exposed to these new SNSs. Hence, the *exclusion* restriction is arguably satisfied.

Meanwhile, usage of "things like Facebook, Instagram, Twitter and Tumblr" in a typical week is intuitively a good indicator of one's SNSs usage in general. Naturally, one might question whether usage of "things like Facebook, Instagram, Twitter and Tumblr"overlaps with SNSs usage to a large extent, invalidating timeSNS as an instrument. To address this concern, two methods of verification are applied. First, a correlation test of timeSNS against SNSs usage generates a coefficient of 0.27, indicating a weak correlation. Second, we cross-tabulate the SNSs dummy with the timeSNS variables. As can be seen in Table 6, the patterns of the timeSNS in the survey week are very similar between those who use SNSs a lot and those who use SNSs not as often. As it turns out, the F-statistic of the first stage estimation increased dramatically after controlling this new variable timeSNS, further justifying the inclusion of this new IV.

4.4 Lewbel method

The Lewbel method is a heteroskedasticity-based instrumental variable method proposed by Lewbel (2012). The method outperforms the conventional IV method in that it does not necessarily require a valid external instrument. Instead, the method can identify endogenous regressors through exploiting the heteroskedasticity in the error term. Additionally, when external instruments are available, the Lewbel method can be used as a robustness check and validity check of the conventional IV method (Baum and Lewbel, 2019). In our case, we utilize both the external instruments, namely the instrumental variables explained in Section 4.3, and the internal instruments which are the entire set of control variables.

Three key assumptions come with the Lewbel method. First, the endogeneity of the instrumented regressor originates from an error component that appears in both stages (i.e. the structural form equation and the reduced form equation). Regarding absenteeism, the error component that can affect SNSs usage and absenteeism simultaneously could be innate ability or discipline (Amez *et al.*, 2019; Baert *et al.*, 2020). Second, any remaining errors are idiosyncratic, providing that the structural model is correctly specified. Given the scarcity of relevant studies in this area, it is difficult to gauge the exact level of appropriateness of the structural model. To specify the structural model as correctly as we can, we attempt to control for an extensive set of covariates in the regression and try alternative specifications in the robustness checks. The third assumption requires that the error term of the reduced form is heteroskedastic, which can be assessed via a Breusch–Pagan test. If satisfied, the third assumption ensures that the constructed instruments are indeed correlated with the endogenous regressor. In sum, the Lewbel method is deemed more appropriate when evidence suggests that these assumptions (should) hold, although the method can still work when the assumptions do not hold (Baum and Lewbel, 2019).

	Freq.	Percent	Freq.	Percent
	Usage on weekdays		Usage at weekend	
Panel A: SNSs = 1				
None	26	1.85	19	1.35
Less than half	532	37.81	362	25.75
About half	425	30.21	416	29.59
More than half	334	23.74	472	33.57
All	90	6.4	137	9.74
Total	1,407	100	1,406	100
Panel B: SNSs = 0				
None	138	17.02	123	15.19
Less than half	427	52.65	345	42.59
About half	145	17.88	186	22.96
More than half	89	10.97	136	16.79
All	12	1.48	20	2.47
Total	811	100	810	100

Table 6.
Distributions of SNSs usage against timeSNS

4.5 Summary of identification strategies

We aim to identify the causal effects as follows. We first try to alleviate reverse causality by lagging explanatory variables [15]. The idea is that absenteeism in a later wave cannot affect SNSs usage in the previous wave. It is more likely that SNSs usage in the previous wave affects absenteeism later. One may argue that there could be unobservables which correlate with both SNSs usage (current or past) and absenteeism at the same time. In the presence of unobservables, the identification may fail (Oster, 2019). In fact, we have adopted a more recent and arguably more flexible technique (Cinelli and Hazlett, 2020) in measuring the magnitude of potential omitted variable bias (OVB). Overall, unobserved confounders (orthogonal to the covariates) that explain more than 1.54–6.24% of the residual variance of both the treatment and the outcome will be strong enough to nullify the results (i.e. bring the point estimate to 0). This is rather strong evidence that OVB is present in the OLS regressions. We thus introduce instrumental variables and apply the conventional IV approach and the heteroskedasticity-based Lewbel approach. In doing so, we identify the causal relationship between SNSs usage and absenteeism.

Additionally, to fully utilize the longitudinal structure of the dataset and to improve the estimation efficiency, we run random effects panel data models assuming random individual specific effects. The SNSs usage variable is recorded consistently in Wave 6 and 7, but not in Wave 8. Thus, our panel data analysis is based on the Wave 6 and 7 data. The two commonly used panel analysis approaches are fixed effects (FE) model and random effects (RE) model. However, compared to RE, FE completely leave the between variation out of the estimation. That means, if the between variation is non-negligible, RE is more appropriate. Using Hausman test (Chi-square: 5.91, p value > 0.8), we cannot reject the null hypothesis that the within effect and the between effect are of similar magnitude. Hence, the random effects approach is more suitable for our purpose, compared to the fixed effects approach. Additionally, we utilize IV method along with the random effects panel data approach to identify causal relationship. The instrument is still whether a teenager more often communicates with close friend face-to-face or electronically.

Understandably, some might be concerned about the omitted variable bias in a random effects model. We thus additionally run a random effects model which allows for both within and between effects (i.e. REWB) (Bell *et al.*, 2019). The REWB model is argued to be superior to both fixed effect models and random effects models since it is the most general of the three and encompasses "all the strengths of the other two" (Bell *et al.*, 2019). In the REWB model, we control for all the covariates in the fixed proportion of the model. In addition, we allow random intercepts at individual level and random slopes by age [16]. The reason that we choose to allow random slopes by age is because teenagers born in different years likely have different SNSs usage habits. Meanwhile, age is a strongly significant variable given our existing specifications. Using the REWB models, we control for fixed effects to the maximum extent while allowing for flexible individual random effects and random slopes.

5. Results

5.1 Effects of SNSs usage on absenteeism

Table 1 presents the effects of SNSs usage on five absenteeism measures. Results in Panel A correspond to the weak hypothesis that only abnormal usage of SNSs hurts one's school attendance. Later analyses will follow the weak hypothesis unless otherwise noted. From the baseline specification in Panel A, we see that, compared to students who use SNSs less often, using SNSs almost every day relates to around 7% points higher probability of a student being late for school, skipping classes, or having trouble not following school rules. All are statistically significant at 1% level. Meanwhile, using SNSs more often also corresponds to higher probability of a student being absent from school with or without parent permission.

Although the significance level changed in several cases, the positive effects are robust to the full specification.

As explained in Section 4.3, to alleviate the endogeneity problem of the treatment variable, IV methods are used. The instrument in the IV model is an indicator variable which equals one if a teenager communicates with close friends more often electronically than face-to-face, and zero otherwise. In Section 4.3, we explain how the instrumental variables arguably satisfy the exclusion restriction and the relevance condition. In Table 1, we provide quantitative evidence by reporting the F-statistic of the first stage estimation along with the estimated coefficients. As can be seen, all the F-statistics are greater than the conventional threshold value of 10 (Staiger and Stock, 1997). Meanwhile, it is possible that means of communication is associated with absenteeism. We thus conducted simple correlation tests between our instrument and the five absenteeism measures, respectively. The correlation coefficients range from 0.02 to 0.09, indicating negligible relationships. We are thus reassured that the means of communication does not directly relate to absenteeism [17]. Additionally, the Durbin tests strongly reject the null hypothesis that the instrumented variable (i.e. SNSs usage) is exogenous in nine out of the ten cases. The IV method unsurprisingly inflates all the coefficients, which is a commonly criticized shortfall of the conventional IV method. Yet, the positive direction of the coefficients remains unchanged, highlighting the causal link between SNSs usage and teenager absenteeism. Despite that our IVs pass the aforementioned validity checks, we are aware that the IV coefficients are inflated heavily and consequently apply a more recent econometric technique (i.e. the Lewbel method) to further investigate the causal relationship. Compared to IV, the Lewbel method is less susceptible to violation of conventional IV assumptions. Because the Lewbel method largely relies on the heteroskedasticity in the error term, Chi-square tests are conducted to confirm that the heteroskedastiticy assumption is satisfied. As can be seen, all results with the Lewbel method remain positive and are close to the OLS estimates. Some results become insignificant probably because of the drop in sample size. Looking through all four specifications, it is unambiguous that, compared to those who use SNSs less frequently, teenagers who use SNSs almost every day is around 8% more likely to be late for school or skip classes.

In parallel, Panel B presents results in line with the strong hypothesis that any kind of SNSs usage hurts attendance. Results are similar to those in Panel A. Interestingly, compared to teenagers who do not use SNSs at all, those who use SNSs are more than 10% are more likely to have trouble not following school rules [18].

5.2 Effect of SNSs usage on study–work choices

Work-related competencies largely affect how smoothly one can find a job. Digital competence is one of the most highly valued work-related competencies nowadays (Oberländer *et al.*, 2020). SNSs usage, as a subset of digital skills, can affect job searching in both ways. On the one hand, it can facilitate job searching by enhancing teenagers' digital competencies and exposing teenagers to personalized job ads. On the other hand, it may impede job searching by making teenagers feel self-content, postpone job searching and delay job entry.

From the results in columns (1) and (2) of Table 7, we see that both abnormal and normal SNSs usage is related to a higher probability of a teenager pausing study at the age of 18. One may wonder whether this relationship is causal. To answer this question, we look at columns (3) and (4). In column (3), SNSs usage is instrumented by the two instrumental variables previously explained: whether one frequently communicates with close friends online or face-to-face and whether one spends no less than half of the time on social media. The F-statistic of the first stage of the IV regression is 54.79, well above the conventional threshold value of 10 for a valid instrument (Staiger and Stock, 1997). We see that those who use SNSs more

Dep. var.: a dummy that equals one for *nostudy* status and zero otherwise	Baseline (1)	Full (2)	IV (3)	Lewbel (4)
SNSs (strong assumption)	0.062** (0.029)	0.061* (0.032)	0.561*** (0.210)	0.048 (0.058)
SNSs (weak assumption)	0.057*** (0.020)	0.066*** (0.023)	0.290*** (0.100)	0.179*** (0.067)
Age		0.069*** (0.022)	0.056** (0.023)	0.062*** (0.023)
Female (yes = 1)		−0.023 (0.024)	−0.061** (0.030)	−0.042 (0.026)
Indigenous people (yes = 1)		−0.014 (0.094)	0.015 (0.097)	0.004 (0.094)
Home language is English (no = 1)		−0.140*** (0.036)	−0.139*** (0.036)	−0.138*** (0.036)
Number of younger siblings		−0.000 (0.012)	0.004 (0.012)	0.001 (0.012)
Number of older siblings		0.002 (0.019)	−0.007 (0.019)	−0.003 (0.019)
Number of same-age siblings		−0.112** (0.052)	−0.122** (0.056)	−0.117** (0.053)
Mother education		−0.012*** (0.004)	−0.011** (0.004)	−0.011*** (0.004)
Peers aged <18 in linked area (%)		0.004 (0.002)	0.004* (0.002)	0.004* (0.002)
IRSAD index		−0.001*** (0.000)	−0.001*** (0.000)	−0.001*** (0.000)
Parent away		0.195* (0.105)	0.190* (0.105)	0.194* (0.104)
Drug use (yes = 1)		0.177*** (0.048)	0.158*** (0.048)	0.167*** (0.048)
Alcohol (more than a few sips = 1)		0.037* (0.022)	0.010 (0.027)	0.024 (0.024)
Boarding school (yes = 1)		−0.007 (0.069)	−0.001 (0.071)	−0.003 (0.069)
School type				
Government school				−0.022 (0.055)
Catholic school		−0.066** (0.028)	−0.076*** (0.029)	−0.094 (0.057)
Independent school		−0.065** (0.028)	−0.058** (0.028)	−0.085 (0.058)
Not in school		0.038 (0.054)	0.017 (0.054)	–
N[a]	2,443	1,801	1,783	1,783
IV F-statistic	–	–	54.79	–
Durbin p-value	–	–	0.015	–
χ^2 of BP test	–	–	–	43.69
State of residence	No	Yes	Yes	Yes
Family income group	No	Yes	Yes	Yes
NAPLAN grades in previous years	No	Yes	Yes	Yes

Note(s): The coefficients of the control variables correspond to the weak assumption. *Parent away* equals one if the main parent was away from the teenager for three or months since the last interview and zero otherwise. [a]The drop in sample sizes in models is due to missing values in the control or instrumental variables. Robust standard errors are in parentheses
***p < 0.01, **p < 0.05, *p < 0.1

Table 7.
Effect of SNSs usage on study-work choices

frequently are around 30% more likely to not study at age 18–19. Column (4) presents results from the heteroskedasticity-based IV method. A Breusch–Pagan test of heteroskedasticity returns a Chi-square statistic of 43.69, strongly rejecting the homoskedasticity assumption. The BP test result suggests that the constructed instruments are indeed correlated with the endogenous regressor, validating the Lewbel method. According to the Lewbel regression, frequent SNSs usage increases the likelihood of a person not studying by around 18%. In sum, although the magnitude of the coefficients varies, the positive sign of the coefficients is robust to four specifications, indicating that frequent SNSs usage increases the probability of a teenager not studying at 18/19 years of age.

We know from Table 3 that the study–work choice variable can be further segregated into four categories: *study–work, study–nowork, nostudy–work* and *nostudy–nowork*. Table 8 shows the multinomial regression results using categorical study–work choice variable. Columns (1) to (3) correspond to the baseline specification. Columns (4) to (6) correspond to the full specification. From the log likelihood ratios, we see that the model with full specification has stronger explanatory power than the baseline model. Unsurprisingly, those who use SNSs more often are more than twice as likely to be in the *nostudy–work* category as those who use SNSs less often. Interestingly though, frequent SNSs users are also more likely to be in the *study–work* category compared to less frequent SNSs users. One possible reason is that frequent access to social media exposes youth to more personalized job opportunities since companies are increasingly using social media as an efficient advertising channel (Kajanová *et al.*, 2017).

Previously, we find that SNSs usage at the age of 14–15 leads to higher probability of a teenager truanting at the age of 16–17. In this section, we find that SNSs usage at the age of 14–15 leads to higher probability of a teenager not studying at the age of 18–19. A natural question is whether the latter outcome (not studying) is a result of the former outcome (truancy). If so, we might have severe omitted variable bias if truancy is uncontrolled when investigating the study–work outcome. Section 5.3 addresses this concern.

5.3 Effect of SNSs usage on study–work choices: after controlling for absenteeism

From the previous section, we see that SNSs usage lead to lower probability of a teenager continuing study after turning 18. The use of SNSs also predicts higher probability of a teenager studying and working at the same time. However, we have also seen that the use of SNSs leads to significantly higher probability of a teenager truanting school in Section 5.1. Is it possible that teenagers' study–work choices are linked to their truanting record, apart from the use of SNSs? To verify whether this is indeed the case, one can try to include the absenteeism measures on the right-hand side of the equation. Given that all five measures are about absenteeism, a natural concern is the multicollinearity problem. To address this concern, we perform the Variance Inflation Factor (VIF) test and find that all the VIF coefficients are around one, detecting no potential multicollinearity.

Then we run the two study–work regressions, the ones in Section 5.2 again but now controlling for the five absenteeism measures in Wave 7. Results are reported in Tables 9 and 10. We see that abnormal SNSs usage increases the likelihood of a teenager being in the *nostudy* status. Although smaller in magnitude, coefficients in the full specification (column (2)) and Lewbel specification (column (4)) remain statistically significant, confirming the robustness of the results in Section 5.2. Interestingly though, after controlling for absenteeism measures, the strong assumption that any kind of SNSs usage decreases the probability of studying is no longer supported by the statistical evidence. Looking at the absenteeism measures in Table 9, three measures remain dormant: being late for school, skipped class and absent with parent permission. Only being absent without parental permission and having trouble not following school rules strongly positively relate to a teenager being in the *nostudy* status. In Table 10, having trouble not following school rules remain strongly significant in all specifications, while being absent without parent permission becomes insignificant in some specifications. Most likely, those who have trouble not following school rules already have a refusal attitude toward attending school and thus stop studying as soon as they can (Kearney, 2008).

5.4 Effect of SNSs usage on absenteeism: panel data random effects approach

Table 11 presents the absenteeism results based on the random effects approach. Basically, abnormal usage of SNSs results in teenagers being 5.7% more likely to be late for school,

Table 8.
Multinomial Logit regression results

	Study-work (1)	NoStudy-work (2)	NoStudy-NoWork (3)	Study-work (4)	NoStudy-work (5)	NoStudy-NoWork (6)
SNSs	1.758*** (0.214)	2.166*** (0.277)	1.093 (0.207)	1.851*** (0.268)	2.476*** (0.400)	1.373 (0.328)
Age				1.052 (0.158)	1.509*** (0.231)	1.112 (0.256)
Female (yes = 1)				1.147 (0.175)	0.954 (0.161)	0.993 (0.248)
Indigenous (yes = 1)				0.419 (0.247)	0.555 (0.306)	0.342 (0.278)
Home language is English (no = 1)				0.631* (0.149)	0.291*** (0.090)	0.482* (0.186)
Number of younger siblings				1.208** (0.100)	1.188* (0.109)	0.990 (0.132)
Number of older siblings				1.274* (0.167)	1.241 (0.172)	1.154 (0.232)
Number of same-age siblings				0.697 (0.252)	0.337** (0.175)	0.595 (0.490)
Mother education				0.922*** (0.023)	0.873*** (0.025)	0.926* (0.043)
Peers aged < 18 in linked area (%)				1.042*** (0.016)	1.053*** (0.019)	1.044 (0.028)
IRSAD index				1.004*** (0.001)	1.000 (0.001)	0.998 (0.002)
Weekend Internet access (min)				1.005 (0.009)	1.008 (0.010)	0.991 (0.015)
Weekdays Internet access (min)				1.000 (0.000)	1.000 (0.000)	1.000 (0.001)
Parent away				0.271 (0.221)	1.064 (0.737)	1.565 (1.292)
Drug use (yes = 1)				2.329* (1.028)	4.868*** (2.215)	2.621* (1.416)
Alcohol (more than a few sips = 1)				1.349** (0.197)	1.464** (0.245)	1.550* (0.361)
Boarding school (yes = 1)				1.321 (0.576)	1.497 (0.728)	0.000*** (0.000)
School type Catholic school				1.994*** (0.359)	1.290 (0.251)	1.045 (0.321)
Independent school				1.388** (0.232)	0.904 (0.175)	1.068 (0.323)
Not in school				1.866 (0.831)	1.618 (0.754)	3.577** (1.808)
Income group FE	Yes	Yes	Yes	Yes	Yes	Yes
City population size FE	Yes	Yes	Yes	Yes	Yes	Yes
State and remoteness FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,292	2,292	2,292	1,806	1,806	1,806
Log-likelihood		−2,775			−1,934	

Note(s): The baseline category is *study-nowork*. The coefficients are relative-risk ratios. *Parent away* equals one if the main parent was away from the teenager for three or months since the last interview and zero otherwise. Robust standard errors clustered by postcode in Wave 7 are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

	Baseline	Full	IV	Lewbel
Dep. var.: a dummy that equals one for nostudy status and zero otherwise				
SNSs (strong assumption)	0.031 (0.030)	0.051 (0.033)	0.562 (0.352)	0.021 (0.053)
SNSs (weak assumption)	0.035 (0.021)	0.059** (0.023)	0.194 (0.120)	0.131* (0.068)
Absenteeism measures				
Late for school	0.013 (0.023)	0.032 (0.025)	0.028 (0.032)	0.030 (0.025)
Skipped class	0.022 (0.030)	−0.026 (0.033)	−0.023 (0.038)	−0.034 (0.033)
Absent with parent permission	−0.025 (0.029)	0.019 (0.031)	−0.002 (0.042)	0.017 (0.032)
Absent without parent permission	0.131*** (0.039)	0.112** (0.044)	0.112** (0.052)	0.118*** (0.044)
Trouble not following school rules	0.087*** (0.022)	0.066*** (0.025)	0.048 (0.032)	0.063** (0.025)
Age		0.070*** (0.022)	0.080*** (0.027)	0.066*** (0.022)
Female (yes = 1)		−0.009 (0.024)	−0.016 (0.033)	−0.021 (0.027)
Indigenous people (yes = 1)		−0.023 (0.097)	−0.084 (0.118)	−0.009 (0.097)
Home language is English (no = 1)		−0.131*** (0.036)	−0.100** (0.047)	−0.128*** (0.036)
No. of younger siblings		0.001 (0.012)	−0.005 (0.016)	0.001 (0.012)
No. of older siblings		−0.000 (0.019)	0.015 (0.024)	−0.003 (0.019)
No. of same-age siblings		−0.107** (0.052)	−0.115 (0.081)	−0.111** (0.052)
Mother education		−0.012*** (0.004)	−0.008 (0.005)	−0.011*** (0.004)
Peers aged <18 in linked area (%)		0.004* (0.002)	0.003 (0.003)	0.005* (0.002)
IRSAD index		−0.001*** (0.000)	−0.000* (0.000)	−0.001*** (0.000)
Parent away		0.211** (0.105)	0.232** (0.105)	0.208** (0.104)
Drug use (yes = 1)		0.156*** (0.049)	0.115** (0.058)	0.152*** (0.048)
Alcohol (more than a few sips = 1)		0.025 (0.023)	0.040 (0.032)	0.019 (0.024)
Boarding school (yes = 1)		−0.008 (0.070)	−0.050 (0.079)	−0.005 (0.069)
School type				
Government school				−0.033 (0.057)
Catholic school		−0.067** (0.029)	−0.054 (0.035)	−0.103* (0.061)
Independent school		−0.066** (0.028)	−0.045 (0.036)	−0.099 (0.061)
Not in school		0.052 (0.056)	0.029 (0.071)	–
N^{a}	2,222	1,785	1,176	1,767
State of residence	No	Yes	Yes	Yes
Family income group	No	Yes	Yes	Yes
NAPLAN grades in previous years	No	Yes	Yes	Yes

Note(s): The coefficients of the control variables correspond to the weak assumption. The absenteeism measures are from Wave 7, in between Wave 6 in which the SNSs usage variable is recorded and Wave 8 in which the study-work choice variable is recorded. 31 *Parent away* equals one if the main parent was away from the teenager for three or months since the last interview and zero otherwise. [a]The drop in sample sizes in models is due to missing values in the control or instrumental variables. Robust standard errors are in parentheses ***p < 0.01, **p < 0.05, *p < 0.1

Table 9.
Effect of SNSs usage on study-work choices

3.1% more likely to skip classes, 3.3% more likely to be absent and 13.6% more likely to have trouble not following school rules. The story is similar when we apply the strong assumption that any SNSs usage is harmful in the panel approach. As for why we consistently observe that SNSs usage results in higher likelihood of absenteeism despite the econometric techniques, many factors could be at play. Most likely though, time spent on SNSs

Table 10. Multinomial logit regression results

	Study-work (1)	NoStudy-work (2)	NoStudy-NoWork (3)	Study-work (4)	NoStudy-work (5)	NoStudy-NoWork (6)
SNSs	1.652*** (0.204)	1.923*** (0.256)	0.986 (0.190)	1.857*** (0.271)	2.394*** (0.392)	1.345 (0.327)
Absenteeism measures	0.961 (0.122)	1.115 (0.166)	0.777 (0.155)	0.947 (0.149)	1.174 (0.211)	1.005 (0.283)
Late for school						
Skip classes	1.087 (0.203)	1.226 (0.247)	0.926 (0.257)	0.977 (0.214)	0.899 (0.222)	0.617 (0.200)
Absent with parent permission	0.970 (0.165)	0.934 (0.173)	0.718 (0.171)	0.931 (0.194)	1.017 (0.236)	1.039 (0.341)
Absent without parent permission	1.261 (0.330)	1.884** (0.492)	2.896*** (0.912)	1.306 (0.446)	1.958* (0.708)	2.756** (1.125)
Trouble not following school rules	1.458*** (0.193)	1.928*** (0.279)	1.867*** (0.368)	1.839*** (0.306)	2.274*** (0.416)	2.038*** (0.520)
Age				1.076 (0.166)	1.577*** (0.244)	1.180 (0.274)
Gender				1.216 (0.184)	1.094 (0.188)	1.128 (0.299)
Indigenous people (yes = 1)				0.434 (0.255)	0.532 (0.301)	0.327 (0.280)
Home language is English (no = 1)				0.690 (0.166)	0.320*** (0.101)	0.500* (0.198)
No. of younger siblings				1.213** (0.102)	1.194* (0.111)	0.981 (0.134)
No. of older siblings				1.272* (0.172)	1.222 (0.175)	1.158 (0.241)
No. of same-age siblings				0.715 (0.267)	0.363* (0.189)	0.593 (0.494)
Mother education				0.923*** (0.024)	0.875*** (0.025)	0.929 (0.043)
Peers aged <18 in linked area (%)				1.043*** (0.016)	1.059*** (0.019)	1.046* (0.027)
IRSAD index				1.004*** (0.001)	1.000 (0.001)	0.998 (0.002)
Parent away				0.423 (0.382)	1.590 (1.270)	2.430 (2.126)
Drug use (yes = 1)				2.302* (1.039)	4.368*** (2.060)	2.494 (1.406)
Alcohol (more than a few sips = 1)				1.304* (0.196)	1.340* (0.233)	1.470 (0.351)
Boarding school (yes = 1)				1.262 (0.565)	1.426 (0.710)	0.000*** (0.000)
School type Catholic school				1.950*** (0.353)	1.241 (0.246)	1.008 (0.315)
Independent school				1.375* (0.231)	0.880 (0.173)	1.070 (0.328)
Not in school				1.657 (0.762)	1.556 (0.744)	3.183** (1.657)
Constant	1.699*** (0.291)	0.768 (0.146)	0.464*** (0.115)	0.005 (0.017)	0.001* (0.005)	3.310 (16.180)
Log-likelihood		−2,654			−1,888	
Income group FE	Yes	Yes	Yes	Yes	Yes	Yes
City population size FE	Yes	Yes	Yes	Yes	Yes	Yes
State and remoteness FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,222	2,222	2,222	1,785	1,785	1,785

Note(s): *Parent away* equals one if the main parent was away from the teenager for three or months since the last interview and zero otherwise. Robust standard errors are in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 11. Absenteeism results: random effects approach

Dep. var.	Late for school (1)	Skip classes (2)	Absent with permission (3)	Absent without permission (4)	Trouble not following rules (5)
Panel A: results with weak hypothesis (abnormal usage is harmful)					
SNSs	0.057*** (0.017)	0.031*** (0.012)	0.033*** (0.012)	0.005 (0.008)	0.136*** (0.017)
N	2,881	2,881	2,882	2,882	2,882
IV	0.270** (0.118)	0.117 (0.080)	0.154* (0.081)	0.070 (0.057)	0.347*** (0.115)
N	2,874	2,874	2,875	2,875	2,875
Panel B: results with strong hypothesis (any usage is harmful)					
SNSs	0.077*** (0.021)	0.015 (0.015)	0.038** (0.015)	0.006 (0.010)	0.134*** (0.021)
N	2,881	2,881	2,882	2,882	2,882
IV	0.376** (0.166)	0.164 (0.112)	0.215* (0.114)	0.097 (0.079)	0.486*** (0.164)
N	2,874	2,874	2,875	2,875	2,875

Note(s): All the dependent variables are dummies which equal 0 for "never" and equal 1 for positive frequencies. Robust standard errors are reported in parentheses

unavoidably substitute teenagers' time commitment to other tasks such as study and sleep. According to recent papers by Amez *et al.* (2021) and Amez *et al.* (2020), smartphone usage is related to deteriorated sleep quality. Although the authors do not find any significant mediating effect of sleep quality on the relationship between smartphone usage and academic performance, it is quite possible that sleep deprivation, partly as a result of SNSs usage, lead to higher probability of absenteeism.

Indeed, we manage to find two sleep-related variables in the dataset and identify interesting patterns [19]. The first variable corresponds to sleep quality and the question reads "During the last month, how well do you feel you have slept in general?" The available answers are "Very well/Fairly well/Fairly badly/Very badly". The second variable relates to whether a teenager has enough sleep or not and the question reads "During the last month, do you think you usually got enough sleep?" The available answers are "Plenty/Just enough/Not quite enough/Not nearly enough". We collapse the two categorical variables into two dummy variables, respectively, so that positive outcomes (i.e. Very well/Fairly well/Plenty/Just enough) equal zero and negative outcomes (i.e. Fairly badly/Very badly/Not quite enough/Not nearly enough) equal one. Then we run multiple linear regressions on the sleep-related variables, controlling for SNSs usage and the full set of covariates as listed in Table 5. Although SNSs usage is not related to the sleep quality indicator ($p = 0.701$), frequent SNSs usage relates to 5.35% ($p = 0.017$) higher probability of a teenager not having enough sleep. Given the significant correlations, we further apply structural equation modeling method to identify whether the *sleep enough* indicator significantly mediates the relationship between absenteeism and SNSs usage. Results are reported in Table 12. In four out of five circumstances, we find that the sleep indicator significantly mediates the effect of SNSs usage on absenteeism. The magnitude ranges from 4.5% to 14.3%. Thus, we have found evidence in line with the conjecture in Amez *et al.* (2021) and Amez *et al.* (2020) that sleep mediates the relationship between SNSs usage and absenteeism.

6. Robustness check

Overall, we have found that SNSs usage results in significantly higher probability of a student truanting school. Despite that we have controlled for an extensive list of covariates, some other factors may still be at play [20]. On the one hand, bullying has been associated with lapses in school attendance (Hutzell and Payne, 2012). We thus try to additionally control

Table 12. Mediating effect of sleep – structural equation modeling

Dep. var.	Late for school (1)	Skip classes (2)	Absent with permission (3)	Absent without permission (4)	Trouble not following rules (5)
Panel A: results with weak hypothesis (only abnormal usage is harmful)					
Direct	0.046* (0.025)	0.085*** (0.023)	0.031 (0.020)	0.038** (0.017)	0.082*** (0.027)
Indirect	0.011*** (0.004)	0.007*** (0.003)	0.004* (0.002)	0.002 (0.002)	0.007** (0.003)
Total	0.058** (0.025)	0.092*** (0.023)	0.035* (0.020)	0.041 (0.017)	0.089*** (0.027)
Proportion of indirect effect	0.190	0.076	0.114	0.049	0.079
Panel B: results with strong hypothesis (any kind of usage is harmful)					
Direct	0.065*** (0.021)	0.082*** (0.190)	0.064*** (0.016)	0.027* (0.014)	0.061*** (0.022)
Indirect	0.011*** (0.003)	0.007*** (0.002)	0.003* (0.002)	0.002 (0.002)	0.007** (0.003)
Total	0.077*** (0.021)	0.089*** (0.019)	0.067*** (0.016)	0.029** (0.014)	0.068*** (0.022)
Proportion of indirect effect	0.143	0.079	0.045	0.069	0.103

Note(s): all the dependent variables are dummies which equal 0 for “never” and equal 1 for positive frequencies. The structural equation model also controls for the set of variables that have been found significant in most specifications (i.e. age, home language, number of same-age siblings, mother education, IRSAD index, drug usage and whether the main parent has been away from home for more than three months) for bully-related variables and re-run Table 1. Results are presented in Table 13. From Table 13, we see that bullying is indeed significant in most specifications, although that mere fact does not change our main story that frequent SNSs usage lead to higher probability of students being absent from school. Bullying is not included in the original specification because many missing values are present and may cause more severe sample selection bias. On the other hand, gaming may also affect teenagers' absenteeism. As can be seen in Table 14 though, gaming rarely appears significant and does not consistently alter the absenteeism results. Neither does it change the SNSs usage coefficient much. Thus, we are reassured of the robustness of our main findings.

To alleviate any concerns regarding potential omitted variable bias in random effects approach, we estimate random effects models allowing for within and between effects (REWB) (Bell *et al.*, 2019). Results of the REWB, along with results from conventional fixed effects models, are reported in Table 15. As expected, Panel A1 and Panel B1 show that the fixed effects results are smaller in scale and not as statistically significant. This is partly because we have lots of fixed effects and consequently end up having fewer degrees of freedom [21]. For bridging purpose, we also run the REWB model only with the fixed effect part before fully allowing between and within variation. Results are shown in Panel A2 and Panel B2 and appear very similar to the full REWB model results. The REWB models include a fixed-effect portion which controls for the full list of covariates as appeared in the full specification and a random-effect portion which, in our case, allows random intercepts at individual level and random slopes by age. From Panel A3 and Panel B3 in Table 15, all

Dep. var.	Late for school (1)	Skip classes (2)	Absent with permission (3)	Absent without permission (4)	Trouble not following rules (5)
Panel A: results with weak hypothesis (only abnormal usage is harmful)					
SNSs	0.032 (0.023)	0.050** (0.021)	0.036* (0.019)	0.004 (0.015)	0.048** (0.024)
Bully	0.103*** (0.023)	0.074*** (0.020)	0.045** (0.018)	0.061*** (0.014)	0.182*** (0.022)
N	1,786	1,786	1,786	1,785	1,788
SNSs	0.864** (0.336)	0.511* (0.264)	0.313 (0.209)	0.360* (0.195)	0.570* (0.299)
Bully	0.030 (0.041)	0.034 (0.031)	0.019 (0.026)	0.031 (0.022)	0.137*** (0.036)
N	1,772	1,772	1,772	1,771	1,774
SNSs	0.177** (0.089)	0.115 (0.079)	0.041 (0.071)	−0.011 (0.055)	0.033 (0.091)
Bully	0.088*** (0.024)	0.067*** (0.020)	0.042** (0.018)	0.063*** (0.014)	0.183*** (0.024)
N	1,772	1,772	1,772	1,771	1,774
Panel B: results with strong hypothesis (any kind of usage is harmful)					
SNSs	0.079** (0.035)	0.071*** (0.026)	0.002 (0.028)	0.026 (0.019)	0.104*** (0.032)
Bully	0.101*** (0.023)	0.074*** (0.020)	0.048*** (0.018)	0.060*** (0.014)	0.180*** (0.022)
N	1,786	1,786	1,786	1,785	1,788
SNSs	1.567** (0.682)	0.927* (0.508)	0.578 (0.408)	0.659* (0.376)	1.044* (0.581)
Bully	0.009 (0.054)	0.022 (0.039)	0.011 (0.031)	0.022 (0.027)	0.124*** (0.044)
N	1,772	1,772	1,772	1,771	1,774
SNSs	0.124** (0.055)	0.073* (0.039)	0.064 (0.047)	0.021 (0.029)	0.093* (0.051)
Bully	0.095*** (0.023)	0.073*** (0.020)	0.042** (0.018)	0.060*** (0.014)	0.181*** (0.023)
N	1,772	1,772	1,772	1,771	1,774

Note(s): All the dependent variables are dummies which equal 0 for "never" and equal 1 for positive frequencies. Full regressions control for demographics as those appear in Table 7. IV stands for the instrumental variable approach estimates. Lewbel stands for the heteroskedasticiy-based IV estimates. Robust standard errors are reported in parentheses. [a]: The drop in sample sizes in models is due to missing values in the control or instrumental variables

Table 13.
Robustness check – absenteeism results

Table 14. Effect of SNSs usage on absenteeism – after controlling for gaming

	Late for school		Skip classes		Absent with permission		Absent without permission		Trouble not following rules	
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: results with weak hypothesis (only abnormal usage is harmful)										
SNSs	0.052*	0.052*	0.069***	0.069***	0.014	0.014	0.024	0.024	0.080***	0.080***
	(0.029)	(0.029)	(0.023)	(0.023)	(0.023)	(0.023)	(0.017)	(0.017)	(0.029)	(0.029)
Gaming		−0.040*		−0.026		−0.030		−0.010		0.009
		(0.023)		(0.020)		(0.018)		(0.015)		(0.024)
N	1,786	1,786	1,786	1,786	1,786	1,786	1,785	1,785	1,788	1,788
Panel B: results with strong hypothesis (any kind of usage is harmful)										
SNSs	0.041*	0.043*	0.057***	0.058***	0.040**	0.042**	0.010	0.011	0.065***	0.065***
	(0.024)	(0.024)	(0.020)	(0.020)	(0.019)	(0.019)	(0.015)	(0.015)	(0.024)	(0.024)
Gaming		−0.042*		−0.028		−0.032*		−0.011		0.007
		(0.023)		(0.020)		(0.019)		(0.015)		(0.024)
N	1,786	1,786	1,786	1,786	1,786	1,786	1,785	1,785	1,788	1,788

Note(s): All the dependent variables are dummies which equal 0 for "never" and equal 1 for positive frequencies. Full regressions control for demographics as those appear in Table 7. Gaming is a dummy which equals one for "almost every day" and zero for "Once or twice a week/A few times a month/Once a month or less/Never"

Dep. var.	Late for school (1)	Skip classes (2)	Absent with permission (3)	Absent without permission (4)	Trouble not following rules (5)
Panel A: results with weak hypothesis (only abnormal usage is harmful)					
Panel A1: fixed effects					
SNSs	0.003 (0.010)	0.025** (0.008)	0.027* (0.012)	0.005 (0.005)	0.044*** (0.008)
N	2,879	2,879	2,880	2,880	2,880
Panel A2: REWB allowing for within effects only					
SNSs	0.060*** (0.008)	0.032*** (0.006)	0.033*** (0.005)	0.004 (0.003)	0.142*** (0.006)
N	3,835	3,834	3,835	3,835	3,837
Panel A3: REWB allowing for within and between effects					
SNSs	0.056*** (0.008)	0.031*** (0.006)	0.033*** (0.004)	0.004 (0.002)	0.138*** (0.006)
N	2,881	2,881	2,882	2,882	2,882
Panel B: results with strong hypothesis (any kind of usage is harmful)					
Panel B1: fixed effects					
SNSs	0.043*** (0.010)	0.009 (0.010)	0.065*** (0.016)	−0.009 (0.007)	0.057*** (0.007)
N	2,879	2,879	2,880	2,880	2,880
Panel B2: REWB allowing for within effects only					
SNSs	0.076*** (0.005)	0.014*** (0.004)	0.038*** (0.008)	0.008*** (0.002)	0.132*** (0.009)
N	3,835	3,834	3,835	3,835	3,837
Panel B3: REWB allowing for within and between effects					
SNSs	0.077*** (0.004)	0.015*** (0.005)	0.041*** (0.008)	0.006*** (0.002)	0.134*** (0.008)
N	2,881	2,881	2,882	2,882	2,882

Note(s): All the dependent variables are dummies which equal 0 for "never" and equal 1 for positive frequencies. Robust standard errors clustered by country of birth are reported in the parentheses

Table 15. Robustness check – random effects model allowing for within and between effects

coefficients remain statistically significant. The magnitudes of the coefficients are also robust to this most flexible specification.

7. Conclusions

Using a longitudinal dataset from Australia, we find that abnormal SNSs usage leads to significantly higher probability of teenagers truanting school. Teenagers who use SNSs abnormally also tend to pause studying after turning 18 years old, although the statistical evidence does not support the same story for those who use SNSs reasonably. Additionally, we identify one important channel through which SNSs usage affect latter study–work choices. That is absenteeism. In other words, increased SNSs usage leads to more frequent absenteeism, which in turn leads to a high probability of a teenager stop studying at the age of 18–19.

This paper contributes to the literature in various aspects. It is the first longitudinal evidence. Meanwhile, a novel set of IV is adopted in this paper. Additionally, novel econometric technique (i.e. Lewbel method) is applied in the analysis. Last but not least, it highlights the importance of controlling for school absenteeism when looking at young adults' labor outcomes. Had absenteeism not been controlled, we would have overestimated the effect of SNS usage on teenagers' study–work choices. Despite the significant effects of SNSs usage on absenteeism, we note that absenteeism is not always a bad thing. Thus, the findings herein are mainly to highlight an undiscovered channel that may affect teenagers' education and labor outcomes. Meanwhile, this paper is not without limitations. First, sample selection issue is present. This is inevitable in longitudinal surveys that span across several years. For example, children in vulnerable environment may be more likely to withdraw from the survey halfway, resulting in vulnerable children being underrepresented in the sample. Second, the IV estimates are inflated heavily compared to the OLS estimates. Although the IVs of choice pass the necessary statistical tests, one should interpret the magnitude of the coefficients attentively. The value of applying the IV method in this paper lies more in its exploratory nature, rather than its quantitative aspect. From a quantitative perspective, the Lewbel estimates are more precise and relatively comparable to the OLS estimates. Last but not least, despite the various techniques we use to alleviate OVB, we are unfortunately unable to quantify the improvement we achieve through these techniques and hence the results may still suffer from OVB.

The school–work choice results have implications for the school-to-work transition situation among Australia youth. It is exciting to see that SNSs usage relates to higher probability of teenagers study and work at the same time. Australian teenagers have always been more likely to work part-time while studying compared to other western countries (Pastore, 2015). Apparently, SNSs usage further enhances this tendency, making Australian teenagers more job-ready compared to similar-age peers around the globe. Meanwhile, SNSs usage also increased the probability of teenagers' work. A natural deduction is that the positive effect of SNSs usage on school-to-work transition outweighs the negative effect. Possibly, SNSs usage enhances teenagers' digital competencies, connects them to wider social network, and promotes personalized job opportunities, all of which facilitate job searching. These results are interesting when compared with the absenteeism results, in the sense that SNSs usage turns out a double-edged sword. On the one hand, it results in higher likelihood of absenteeism at school. On the other hand, it presumably increases the likelihood of teenagers having smoother school-to-work transition. We note that student employment is found to decrease the probability of a student continuing education or enter tertiary education (Neyt *et al.*, 2019). Given the observed dynamic patterns, the results on the school–work choices should not be extended beyond its scope. The medium to long run impact of SNSs usage on youth study–work outcomes is left to future research.

The findings in this paper have both practical and social implications. On a practical level, stakeholders, especially parents and school teachers, can reflect on teenagers' social media usage patterns. Parents can utilize digital safety app to restrict children's overuse of social media apps (e.g. by setting a time limit). Teachers can remind students to be more mindful of their social media engagement out of class. From the social perspective, social media giants could take greater social responsibility and provide restricted access to younger users, although these actions might be at the cost of their short-run profit. Correspondingly, the government could compensate these companies by reducing tax or providing subsidies.

Additionally, the government could promote more diverse after-class entertainment options, so that children are less attracted by social media and hence less likely to overuse it.

Notes

1. The data can be accessed via this link
2. Situations vary slightly across states. See Study Australia for more information.
3. Accessible via this link
4. Note SNSs usage is distinct from Internet usage in general, as can be validated by the following observations from the dataset. In Wave 7, questions are asked regarding the proportion of Internet access time spent on SNSs during weekdays and weekends. On weekdays, more than 50% of the surveyed teenagers report spending less than half or none of the Internet time on SNSs, while 4.55% of the respondents report spending all the Internet time on SNSs. At weekends, around 38% respondents report spending less than half or none of the Internet time on SNSs and around 27% report spending roughly half of the Internet time on SNSs, although only 6.9% report spending all the Internet time on SNSs. Hence, SNSs usage patterns are quite distinct from with Internet usage patterns.
5. As mentioned in Section 3.2, the threshold age is smaller than 18 in some states.

6. See here for more information.
7. The inaccuracy in the SNSs usage variable could have been problematic if, for example, one studied the effect of gaming, because it would be very difficult to identify the SNSs usage that are dedicated to gaming. Yet, it is less of a concern since we study SNSs usage itself.
8. See here for more information.
9. See here
10. The mapping is prepared by the OECD, accessible via this link
11. If Certificate III and Certificate IV is not differentiated in a record, then their average (i.e. 13) is used.
12. The Internet quality across different geographical areas would have been great as an instrument. Yet, in our dataset, all the statistical areas are de-identified and presented as pseudo values. Thus, it is not possible for us to match the Internet quality in different areas with their respective statistical areas.
13. The data can be accessed via this link
14. See this link
15. An objection to the use of lagged variables is as follows: Consider an equation that looks like this: $Y(t) = X(t-1)$. If $X(t)$ and $X(t-1)$ are correlated, then $X(t-1)$ and $Y(t)$ will be endogenous. We thank an anonymous reviewer for this caveat.
16. We have tried varying slopes across various other variables. The results remain virtually the same.
17. We thank an anonymous reviewer for pointing out this caveat.
18. Absenteeism results by gender are also explored but not presented here since the patterns do not differ much between different genders.
19. We greatly thank an anonymous reviewer for recommending this exploration.
20. We thank two anonymous reviewers for suggesting bullying, gaming and other potentially interesting variables (i.e. insufficient teacher involvement and different interests concerning the subjects taught). The other variables are unfortunately not available in our dataset and may be considered by future researchers.
21. We thank an anonymous reviewer for correcting an imprecise statement here.

References

Allam, H., Blustein, J., Bliemel, M. and Spiteri, L. (2012), "Knowledge contribution in social media: exploring factors influencing social taggers' acceptance towards contributing and sharing tags", in *International Conference on Information Systems, Technology and Management*, Springer, pp. 112-123.

Amez, S. and Baert, S. (2020), "Smartphone use and academic performance: a literature review", *International Journal of Educational Research*, Vol. 103, 101618.

Amez, S., Vujić, S., De Marez, L. and Baert, S. (2019), "Smartphone Use and Academic Performance: First Evidence from Longitudinal Data", *New Media and Society*, doi: 10.1177/146144482110123 74.

Amez, S., Vujić, S., Soffers, P. and Baert, S. (2020), "Yawning while scrolling? Examining gender differences in the association between smartphone use and sleep quality", *Journal of Sleep Research*, Vol. 29 No. 6, e12971.

Amez, S., Vujić, S., Abrath, M. and Baert, S. (2021), *Deteriorated Sleep Quality Does Not Explain the Negative Impact of Smartphone Use on Academic Performance*, Technical report, GLO Discussion Paper (No. 879).

Andreassen, C.S. (2015), “Online social network site addiction: a comprehensive review”, *Current Addiction Reports*, Vol. 2 No. 2, pp. 175-184.

Austin, W.A. and Totaro, M.W. (2011), “Gender differences in the effects of internet usage on high school absenteeism”, *The Journal of Socio-Economics*, Vol. 40 No. 2, pp. 192-198.

Australian Bureau Of Statistics, A. (2011), *Socio-economic Indexes for Areas (Seifa)*, Australian Bureau of Statistics, Canberra.

Baert, S., Vujić, S., Amez, S., Claeskens, M., Daman, T., Maeckelberghe, A., Omey, E. and De Marez, L. (2020), “Smartphone use and academic performance: correlation or causal relationship?”, *Kyklos*, Vol. 73 No. 1, pp. 22-46.

Baum, C.F. and Lewbel, A. (2019), “Advice on using heteroskedasticity-based identification”, *The Stata Journal*, Vol. 19 No. 4, pp. 757-767.

Bell, A., Fairbrother, M. and Jones, K. (2019), “Fixed and random effects models: making an informed choice”, *Quality and Quantity*, Vol. 53 No. 2, pp. 1051-1074.

Boyd, D.M. and Ellison, N.B. (2007), “Social network sites: definition, history, and scholarship”, *Journal of Computer-Mediated Communication*, Vol. 13 No. 1, pp. 210-230.

Brandtzæg, P.B. (2012), “Social networking sites: their users and social implications—a longitudinal study”, *Journal of Computer-Mediated Communication*, Vol. 17 No. 4, pp. 467-488.

Cinelli, C. and Hazlett, C. (2020), “Making sense of sensitivity: extending omitted variable bias”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 82 No. 1, pp. 39-67.

Clarke, P.M., Fiebig, D.G. and Gerdtham, U.-G. (2008), “Optimal recall length in survey design”, *Journal of Health Economics*, Vol. 27 No. 5, pp. 1275-1284.

Durak, H.Y. (2020), “Modeling of variables related to problematic internet usage and problematic social media usage in adolescents”, *Current Psychology*, Vol. 39 No. 4, pp. 1375-1387.

Hutzell, K.L. and Payne, A.A. (2012), “The impact of bullying victimization on school avoidance”, *Youth Violence and Juvenile Justice*, Vol. 10 No. 4, pp. 370-385.

Kajanová, H., Sedlácek, M. and Soósová, V. (2017), “Attitudes of young people to job searching through social media: case of Slovakia”, *Economics and Sociology*, Vol. 10 No. 1, p. 152.

Kearney, C.A. (2008), “School absenteeism and school refusal behavior in youth: a contemporary review”, *Clinical Psychology Review*, Vol. 28 No. 3, pp. 451-471.

Keles, B., McCrae, N. and Grealish, A. (2020), “A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents”, *International Journal of Adolescence and Youth*, Vol. 25 No. 1, pp. 79-93.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T. and Scherlis, W. (1998), “Internet paradox: a social technology that reduces social involvement and psychological well-being?”, *American Psychologist*, Vol. 53 No. 9, p. 1017.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V. and Crawford, A. (2002), “Internet paradox revisited”, *Journal of Social Issues*, Vol. 58 No. 1, pp. 49-74.

Kreski, N., Platt, J., Rutherford, C., Olfson, M., Odgers, C., Schulenberg, J. and Keyes, K.M. (2021), “Social media use and depressive symptoms among United States adolescents”, *Journal of Adolescent Health*, Vol. 68 No. 3, pp. 572-579.

Lenhart, A., Smith, A., Anderson, M., Duggan, M. and Perrin, A. (2015), *Teens, Technology and Friendship*, Vol. 10, Pew Research Center, Washington, DC.

Lewbel, A. (2012), “Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models”, *Journal of Business and Economic Statistics*, Vol. 30 No. 1, pp. 67-80.

Lim, S.S. (2013), *Young People Sharing Norms and Collective Identities with and through Media*, The Routledge International Handbook of Children, Adolescents, and Media, London, pp. 322-328.

McDool, E., Powell, P., Roberts, J. and Taylor, K. (2020), “The internet and children’s psychological wellbeing”, *Journal of Health Economics*, Vol. 69, 102274.

Neyt, B., Omey, E., Verhaest, D. and Baert, S. (2019), "Does student work really affect educational outcomes? A review of the literature", *Journal of Economic Surveys*, Vol. 33 No. 3, pp. 896-921.

Obar, J.A. and Wildman, S.S. (2015), "Social media definition and the governance challenge an introduction to the special issue", *Telecommunications Policy*, Vol. 39 No. 9, pp. 745-750.

Oberländer, M., Beinicke, A. and Bipp, T. (2020), "Digital competencies: a review of the literature and applications in the workplace", *Computers and Education*, Vol. 146, 103752.

Oster, E. (2019), "Unobservable selection and coefficient stability: theory and evidence", *Journal of Business and Economic Statistics*.

Ostic, D., Qalati, S.A., Barbosa, B., Shah, S.M.M., Galvan Vela, E., Herzallah, A.M. and Liu, F. (2021), "Effects of social media use on psychological well-being: a mediated model", *Frontiers in Psychology*, Vol. 12, p. 2381.

Parker, P.D., Thoemmes, F., Duineveld, J.J. and Salmela-Aro, K. (2015), "I wish I had (not) taken a gap-year? The psychological and attainment outcomes of different post-school pathways", *Developmental Psychology*, Vol. 51 No. 3, p. 323.

Pastore, F. (2015), *The Youth Experience Gap: Explaining National Differences in the School-to-Work Transition*, Springer Heidelberg, Physica Verlag.

Pastore, F. (2018), "Why is youth unemployment so high and different across countries?", *IZA World of Labor 420*, doi: 10.15185/izawol.420.

Pastore, F. and Zimmermann, K.F. (2019), "Understanding school-to-work transitions", *International Journal of Manpower*, Vol. 40 No. 3, pp. 374-378.

Pastore, F., Quintano, C. and Rocca, A. (2020), "Stuck at a crossroads? The duration of the Italian school-to-work transition", *International Journal of Manpower*, Vol. 42 No. 3, pp. 442-469.

Pastore, F., Quintano, C. and Rocca, A. (2021), "Some young people have all the luck! the duration dependence of the school-to-work transition in Europe", *Labour Economics*, Vol. 70, 101982.

Posso, A. (2016), "Internet usage and educational outcomes among 15-year old Australian students", *International Journal of Communication*, Vol. 10, p. 26.

Rose Birch, E. and Miller, P.W. (2007), "The characteristics of 'gap-year' students and their tertiary academic outcomes", *Economic Record*, Vol. 83 No. 262, pp. 329-344.

Staiger, D. and Stock, J.H. (1997), "Instrumental variables regression with weak instruments", *Econometrica: Journal of the Econometric Society*, Vol. 65, pp. 557-586.

Swist, T., Collin, P., McCormack, J. and Third, A. (2015), *Social Media and the Wellbeing of Children and Young People: A Literature Review*, Commissioner for Children and Young People, Western Australia.

Young, K.S. and Case, C.J. (2004), "Internet abuse in the workplace: new trends in risk management", *Cyber Psychology and Behavior*, Vol. 7 No. 1, pp. 105-111.

Corresponding author
Jie Chen can be contacted at: jiechenunsw@outlook.com