

Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies



Amy Orben¹ and Andrew K. Przybylski^{1,2}

¹Department of Experimental Psychology, University of Oxford, and ²Oxford Internet Institute, University of Oxford

Psychological Science
1–15

© The Author(s) 2019



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0956797619830329

www.psychologicalscience.org/PS



Abstract

The notion that digital-screen engagement decreases adolescent well-being has become a recurring feature in public, political, and scientific conversation. The current level of psychological evidence, however, is far removed from the certainty voiced by many commentators. There is little clear-cut evidence that screen time decreases adolescent well-being, and most psychological results are based on single-country, exploratory studies that rely on inaccurate but popular self-report measures of digital-screen engagement. In this study, which encompassed three nationally representative large-scale data sets from Ireland, the United States, and the United Kingdom ($N = 17,247$ after data exclusions) and included time-use-diary measures of digital-screen engagement, we used both exploratory and confirmatory study designs to introduce methodological and analytical improvements to a growing psychological research area. We found little evidence for substantial negative associations between digital-screen engagement—measured throughout the day or particularly before bedtime—and adolescent well-being.

Keywords

large-scale social data, digital technology use, adolescents, well-being, time-use diary, specification-curve analysis, open materials, preregistered

Received 6/27/18; Revision accepted 11/27/18

As digital screens become an increasingly integral part of daily life for many, concerns about their use have become common (see Bell, Bishop, & Przybylski, 2015, for a review). Scientists, practitioners, and policymakers are now looking for evidence that could inform possible large-scale interventions designed to curb the suspected negative effects of excessive adolescent digital engagement (UK Commons Select Committee, 2017). Yet there is still little consensus as to whether and, if so, how digital-screen engagement affects psychological well-being; results of studies have been mixed and inconclusive, and associations—when found—are often small (Etchells, Gage, Rutherford, & Munafò, 2016; Orben & Przybylski, 2019; Parkes, Sweeting, Wight, & Henderson, 2013; Przybylski & Weinstein, 2017; Smith, Ferguson, & Beaver, 2018).

In most previous work, researchers considered the amount of time spent using digital devices, or *screen time*, as the primary determinant of positive or negative technology effects (Neuman, 1988; Przybylski & Weinstein,

2017). It is therefore imperative that such work incorporates high-quality assessments of screen time. Yet with the vast majority of studies relying on retrospective self-report scales, research indicates that there is good reason to believe that current screen-time measurements are lacking in quality (Scharnow, 2016). On the one hand, people are not skilled at perceiving the time they spend engaging in specific activities (Grondin, 2010). On the other hand, there are also a myriad of additional reasons why people fail to give accurate retrospective self-report judgments (e.g., Boase & Ling, 2013; Schwarz & Oyserman, 2001).

Recent work has demonstrated that only one third of participants provide accurate judgments when asked about their weekly Internet use, while 42% overestimate

Corresponding Author:

Amy Orben, University of Oxford, Department of Experimental Psychology, New Radcliffe House, Oxford OX2 6GG England

E-mail: amy.orben@psy.ox.ac.uk

and 26% underestimate their usage (Scharkow, 2016). Inaccuracies vary systematically as a function of actual digital engagement (Vanden Abeele, Beullens, & Roe, 2013; Wonneberger & Irazoqui, 2017): Heavy Internet users tend to underestimate the amount of time they spend online, while infrequent users overreport this behavior (Scharkow, 2016). Both these trends have been replicated in subsequent studies (Araujo, Wonneberger, Neijens, & de Vreese, 2017). There are therefore substantial and endemic issues regarding the majority of current research investigating digital-technology use and its effects.

Direct tracking of screen time and digital activities on the device level is a promising approach for addressing this measurement problem (Andrews, Ellis, Shaw, & Piwek, 2015; David, Roberts, & Christenson, 2018), yet the method comes with technical issues (Miller, 2012) and is still limited to small samples (Junco, 2013). Given the importance of rapidly gauging the impact of screen time on well-being, other approaches for measuring the phenomena—approaches that can be implemented more widely—are needed for psychological science to progress.

To this end, a handful of recent studies have applied experience-sampling methodology, asking participants specific technology-related questions throughout the day (Verduyn et al., 2015) or after specific bouts of digital engagement (Masur, 2018). This method is complemented by studies using time-use diaries, which require participants to recall what activities they were engaged in during prespecified days; this approach builds a detailed picture of the participants' daily life (Hanson, Drumheller, Mallard, McKee, & Schlegel, 2010). Because most time-use diaries ask participants to recount small time windows (e.g., every 10 min), they facilitate the summation of total time spent engaging with digital screens and allow for investigation into the time of day that these activities occur. Time-use diaries could therefore extend and complement the more commonly used self-report measurement methodology. Yet work using these promising time-use-diary measures has focused mainly on single smaller data sets, has not been preregistered, and has not examined the effect of digital engagement on psychological well-being.

Specifically, time-use diaries allow us to examine how digital-technology use before bedtime affects both sleep quality and duration. Researchers have postulated that by promoting users' continued availability and fear of missing out, social media platforms can decrease the amount of time adolescents sleep (Scott, Biello, & Cleland, 2018). Previous research found negative effects when adolescents engage with digital screens 30 min (Levenson, Shensa, Sidani, Colditz, & Primack, 2017), 1 hr (Harbard, Allen, Trinder, & Bei, 2016), and 2 hr (Orzech, Grandner, Roane, & Carskadon, 2016) before bedtime. This could be attributable to delayed bedtimes (Cain & Gradisar,

2010; Orzech et al., 2016) or to difficulties in relaxing after engaging in stimulating technology use (Harbard et al., 2016).

The Present Research

In this research, we focused on the relations between digital engagement and psychological well-being using both time-use diaries and retrospective self-report data obtained from adolescents of three different countries—Ireland, the United States, and the United Kingdom. Across all data sets, our aim was to determine the direction, magnitude, and statistical significance of these relations, with a particular focus on the effects of digital engagement before bedtime. In order to clarify the mixed literature and provide high generalizability and transparency, we used the first two studies to extend a general research question concerning the link between screen time and well-being into specific hypotheses. These theories were then tested in a third confirmatory study. More specifically, we used specification-curve analysis (SCA) to identify promising links in our two exploratory studies, generating informed data- and theory-based hypotheses. The robustness of these hypotheses were then evaluated in a third study using a preregistered confirmatory design. By subjecting the results from the first two studies to the highest methodological standards of testing, we aimed to shed further light on whether digital engagement has reliable, measurable, and substantial associations with the psychological well-being of young people.

Exploratory Studies

Method

Data sets and participants. Data from two nationally representative data sets collected in Ireland and the United States were used to explore the plausible links between psychological well-being and digital engagement, generating hypotheses for subsequent testing. We selected both data sets because they were large in comparison with normal social-psychological-research data sets (total $N = 5,363$; Ireland: $n = 4,573$, United States: $n = 790$ after data exclusions); they were also nationally representative and had open and harmonized well-being and time-use-diary measurements. Because technology use changes so rapidly, only the most recent wave of time-use diaries was analyzed so that the data would reflect the current state of digital engagement.

The first data set under analysis was Growing Up in Ireland (GUI; Williams et al., 2009). In our study, we focused on the GUI child cohort that tracked 5,023 nine-year olds, recruited via random sampling of primary schools. The wave of interest took place between August

2011 and March 2012 and included 2,514 boys and 2,509 girls, mostly aged thirteen (4,943 thirteen-year-olds, 24 twelve-year-olds, and 56 fourteen-year-olds). The time-use diaries were completed on a day individually designated by the head office (either weekend or weekday) after the primary interview of both children and their caretakers. After data exclusions, 4,573 adolescents were included in the study.

Collected between 2014 and 2015, the second data set of interest was the United States Panel Study of Income Dynamics (PSID; Survey Research Center, Institute for Social Research, University of Michigan, 2018), which included 741 girls and 767 boys. It encompassed participants from a variety of age groups: 108 eight-year-olds, 100 nine-year-olds, 110 ten-year-olds, 89 eleven-year-olds, 201 twelve-year-olds, 213 thirteen-year-olds, 190 fourteen-year-olds, 186 fifteen-year-olds, 165 sixteen-year-olds, 127 seventeen-year-olds, and 19 who did not provide an age. We selected only those 790 participants who were between the ages of 12 and 15, to match the age ranges in the other data sets used. The sample was collected by involving all children in households already interviewed by the PSID who were descended from either the original families recruited in 1968 or the new immigrant family sample added in 1997. Those participants in the child supplement who were selected to receive an in-home visit were asked to complete two time-use diaries on randomly assigned days (one on a weekday and one on a weekend day).

Ethical review. The Research Ethics Committee of the Health Research Board in Ireland gave ethical approval to the GUI study. The University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board reviews the PSID annually to ensure its compliance with ethical standards.

Measures. This research examined a variety of well-being and digital-screen-engagement measures. While each data set included a range of well-being questionnaires, we considered only those measures present in at least one of the exploratory data sets (i.e., the GUI and the PSID) and in the data set used for our confirmatory study (detailed below). Thus, measures included the popular Strengths and Difficulties Questionnaire (SDQ) completed by caretakers (part of the GUI and the confirmatory study) and two well-being questionnaires filled out by adolescents—the Short Mood and Feelings Questionnaire (in the GUI and the confirmatory study) and the Children’s Depression Inventory in the PSID. In addition, we relied on the Rosenberg Self-Esteem Scale (part of the PSID and the confirmatory study).

Adolescent well-being. The first measure of adolescent well-being considered was the SDQ completed by the

Irish participants’ primary caretakers (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000). This measure of psychosocial functioning has been widely used and validated in school, home, and clinical contexts. It includes 25 questions, 5 each about prosocial behavior, hyperactivity or inattention, emotional symptoms, conduct problems, and peer-relationship problems (0 = *not true*, 1 = *somewhat true*, 2 = *certainly true*; prosocial behavior was not included in our analyses, and the scale was subsequently reverse scored; see the Supplemental Material available online).

The second measure of adolescent well-being was an abbreviated version of the Rosenberg Self-Esteem Scale completed by U.S. participants (Robins, Hendin, & Trzesniewski, 2001). This was a five-item measure that asked, “How much do you agree or disagree with the following statement?” Answer choices were, “On the whole, I am satisfied with myself”; “I feel like I have a number of good qualities”; “I am able to do things as well as most other people”; “I am a person of value”; and “I feel good about myself.” The participants answered these questions on a four-item Likert scale that ranged from *strongly disagree* (1) to *strongly agree* (4).

Third, for the Irish data set, we included the Child Depression Inventory as a negative indicator of well-being. The adolescent participants answered questions about how they felt or acted in the past 2 weeks using a three-level Likert scale that ranged from *true* to *not true*. Items included “I felt miserable or unhappy,” “I didn’t enjoy anything at all,” “I felt so tired I just sat around and did nothing,” “I was very restless,” “I felt I was no good any more,” “I cried a lot,” “I found it hard to think properly or concentrate,” “I hated myself,” “I was a bad person,” “I felt lonely,” “I thought nobody really loved me,” “I thought I could never be as good as other kids,” and “I did everything wrong.” We subsequently reverse-scored items so they instead measured adolescent well-being.

Finally, for the U.S. sample, we included the Short Mood and Feelings Questionnaire as a measure of adolescent well-being. The participants were asked to think about the last 2 weeks and select a sentence that best described their feelings (see the Supplemental Material for the sentences used). The choices are very similar to those in the 12 questions about subjective affective states and general mood asked in the confirmatory data set detailed later.

Adolescent digital engagement. The study included two varieties of digital-engagement measures: retrospective self-report measures of digital engagement and estimates derived from time-use diaries. Details regarding these measures varied for each data set because of differences in the questionnaires and time diaries used. For all data sets, we removed participants who filled out a

time-use diary during a weekday that was not term or school time, and if participants went to bed after midnight (after the time-use diary was concluded), we coded them as going to bed at midnight.

The Irish data set included three questions asking participants to think of a normal weekday during term time and estimate, “How many hours do you spend watching television, videos or DVDs?” “How much time do you spend using the computer (do not include time spent using computers in school)?” and “How much time do you spend playing video games such as PlayStation, Xbox, Nintendo, etc.?” Participants could answer in hours and minutes, but responses were recoded by the survey administrator into a 13-level scale.¹ We took the mean of these measures to obtain a general digital-engagement measure. In the U.S. data set, adolescents were asked, “In the past 30 days, how often did you use a computer or other electronic device (such as a tablet or smartphone)” for any of the following: “for school work done at school or at home,” “for these types of online activities (visiting a newspaper or news-related website; watch or listen to music, videos, TV shows, or movies; follow topics or people that interest you on websites, blogs, or social media sites (like Facebook, Instagram or Twitter), not including following or interacting with friends or family online),” “to play games,” and “for interacting with others.” Participants answered using a 5-point Likert scale ranging from *never* (1) to *every day* (5). For the U.S. data, we took the mean of these four items to obtain a general digital-engagement measure.

The study focused on five discrete measures from the participants’ self-completed time-use diaries: (a) whether the participants reported engaging with any digital screens, (b) how much time they spent doing so, and whether they did so (c) 2 hr, (d) 1 hr, and (e) 30 min before going to bed. We separated these numerical measures for weekend and weekday, resulting in a total of 10 different variables.

Each time-use diary, although harmonized by study administrators, was administered and coded slightly differently. The Irish data set contained 21 pre-coded activities that participants could select for each 15-min period. These included the four categories we then aggregated into our digital-engagement measure: “using the internet/emailing (including social networking, browsing etc.),” “playing computer games (e.g., PlayStation, PSP, X-Box or Wii),” “talking on the phone or texting,” or “watching TV, films, videos or DVDs.” In the U.S. data set, participants (or their caretakers) could report their activities freely, including primary and secondary activities, duration, and where the activity occurred. Research assistants coded these activities afterward. There were 13 codes aggregated in our

digital-engagement measure, including lessons in using a computer or other electronic device, playing electronic games, other technology-based recreational activities, communication using technology/social media, texting, uploading or creating Internet content, nonspecific work with technology such as installing software or hardware, photographic processing, and other activities involving a computer or electronic device. We aggregated these measures and did not include them in our analyses separately because there were too few people who scored on any one coded variable.

Time-use-diary measures commonly have high positive skew: Many participants do not note down the activity at all, while only a few report spending much time on the activity. It is common practice to address this by splitting the time-use variable into two measures, with the first reflecting participation and the second reflecting amount of participation (i.e., the time spent doing this activity; Hammer, 2012; Rohrer & Lucas, 2018). *Participation* is a dichotomous variable representing whether a participant reported engaging in the activity on a given day; *time spent* is a continuous variable that represents the amount of engagement for participants who reported doing the activity.

In addition to including these two different measures—separately for weekends and weekdays—we also created six measures to assess technology use before bedtime. These measures were dichotomous, simply indicating whether the participant had used technology in the specified time interval. These time intervals were 30 min, 1 hr, and 2 hr before bedtime, assessed separately on a weekend day and a weekday.

Covariate and confounding variables. Minimal covariates were incorporated in these exploratory analyses—just gender and age for both Irish and U.S. data sets—to prevent spurious correlations or conditional associations from complicating our hypothesis-generating process.

Analytic approach

To examine the correlation between technology use and well-being, we used an SCA approach proposed by Simonsohn, Simmons, and Nelson (2015) and applied in recent articles by Rohrer, Egloff, and Schmukle (2017) and Orben and Przybylski (2019). SCA enables researchers to implement many possible analytical pathways and interpret them as one entity, respecting that the “garden of forking paths” allows for many different data-analysis options which should be taken into account in scientific reporting (Gelman & Loken, 2013). Because the aim for these analyses was to generate informed data- and theory-driven hypotheses to then test in a later confirmatory study, the analyses consisted of four steps.

Table 1. Specifications Tested in the Irish, American, and British Data Sets

Decision	Hypothesis-generating studies		Hypothesis-testing study
	Ireland	United States	United Kingdom
Operationalizing adolescent well-being	Strengths and Difficulties Questionnaire; well-being; Child Depression Inventory	Well-being: Short Moods and Feelings Questionnaire; Rosenberg Self-Esteem Scale	Strengths and Difficulties Questionnaire; well-being; Short Moods and Feelings Questionnaire; Rosenberg Self-Esteem Scale
Operationalizing digital engagement	Retrospective self-report measure; time-use-diary measures (weekday and weekend separately): (a) participation, (b) time spent, (c) < 2 hr before bedtime, (d) < 1 hr before bedtime, and (e) < 30 min before bedtime	Retrospective self-report measure; time-use-diary measures (weekday and weekend separately): (a) participation, (b) time spent, (c) < 2 hr before bedtime, (d) < 1 hr before bedtime, and (e) < 30 min before bedtime	Retrospective self-report measure; time-use-diary measures (weekday and weekend separately): (a) participation, (b) time spent, (c) < 2 hr before bedtime, (d) < 1 hr before bedtime, and (e) < 30 min before bedtime
Inclusion of control variables	No controls; all controls	No controls; all controls	No controls; all controls

Correlations between retrospective reports and time-use-diary estimates. The first analytical step was to examine the correlations between retrospective self-report and time-use-diary measures of digital engagement, to gauge whether they were measuring similar or removed concepts. This was done to inform later interpretations of the SCA and to give valuable insights to researchers about such widely used measures.

Identifying specifications. We then decided which theoretically defensible specifications to include in the SCA. While this was done a priori for all studies, it was specifically preregistered only for the confirmatory study. The three main analytical choices addressed in the SCA were how to measure well-being, how to measure digital engagement, and whether to include statistical controls or not (see Table 1). Three different possible measures of well-being were included in the exploratory data sets: the SDQ, the reversed Child Depression Inventory or Short Moods and Feelings Questionnaire, and the Rosenberg Self-Esteem Scale. There were 11 possible measures of digital engagement, including the retrospective self-report measure and the time-use-diary measures separated for weekend day or weekday (participation, time spent, and engagement at < 2 hr, < 1 hr, and < 30 min before bedtime). Lastly, there was a choice of whether to include controls in the subsequent analyses or not.

Implementing specifications. Taking each specification in turn, we ran a linear regression to obtain the standardized regression coefficient linking digital engagement measurements to well-being outcomes. To do so, we first used the various digital-engagement measures to predict the specific well-being questionnaires identified in the

study. The regression either did or did not include covariates, depending on the specifications. We noted the standardized regression coefficient, the corresponding p value, and the partial r^2 . We also ran 500 bootstrapped models of each SCA to obtain the 95% confidence intervals (CIs) around the standardized regression coefficient and the effect-size measure.

The specifications were then ranked by their regression coefficient and plotted in a specification curve, where the spread of the associations is most clearly visualized. The bottom panel of the specification-curve plot illustrates what analytical decisions lead to what results, creating a tool for mapping out the too-often-invisible garden of forking paths (for an example, see Fig. 1).

Statistical inferences. Bootstrapped models were implemented to examine whether the associations evident in the calculated specifications were significant (Orben & Przybylski, 2019; Simonsohn et al., 2015). We were particularly interested in the different measures and the timing of digital-technology use, so we ran a separate significance test for each technology-use measure. Our bootstrapped approach was necessary because the specifications do not meet the independence assumption of conventional statistical testing. We created data sets in which we knew the null hypothesis was true and examined the median point estimate (measured using the median regression coefficient) and number of significant specifications in the dominant direction (the sign of the majority of the specifications) they produced. We used these two significance measures, as proposed by Simonsohn and colleagues, but do not report the number of specifications in the dominant direction—a significance measure also

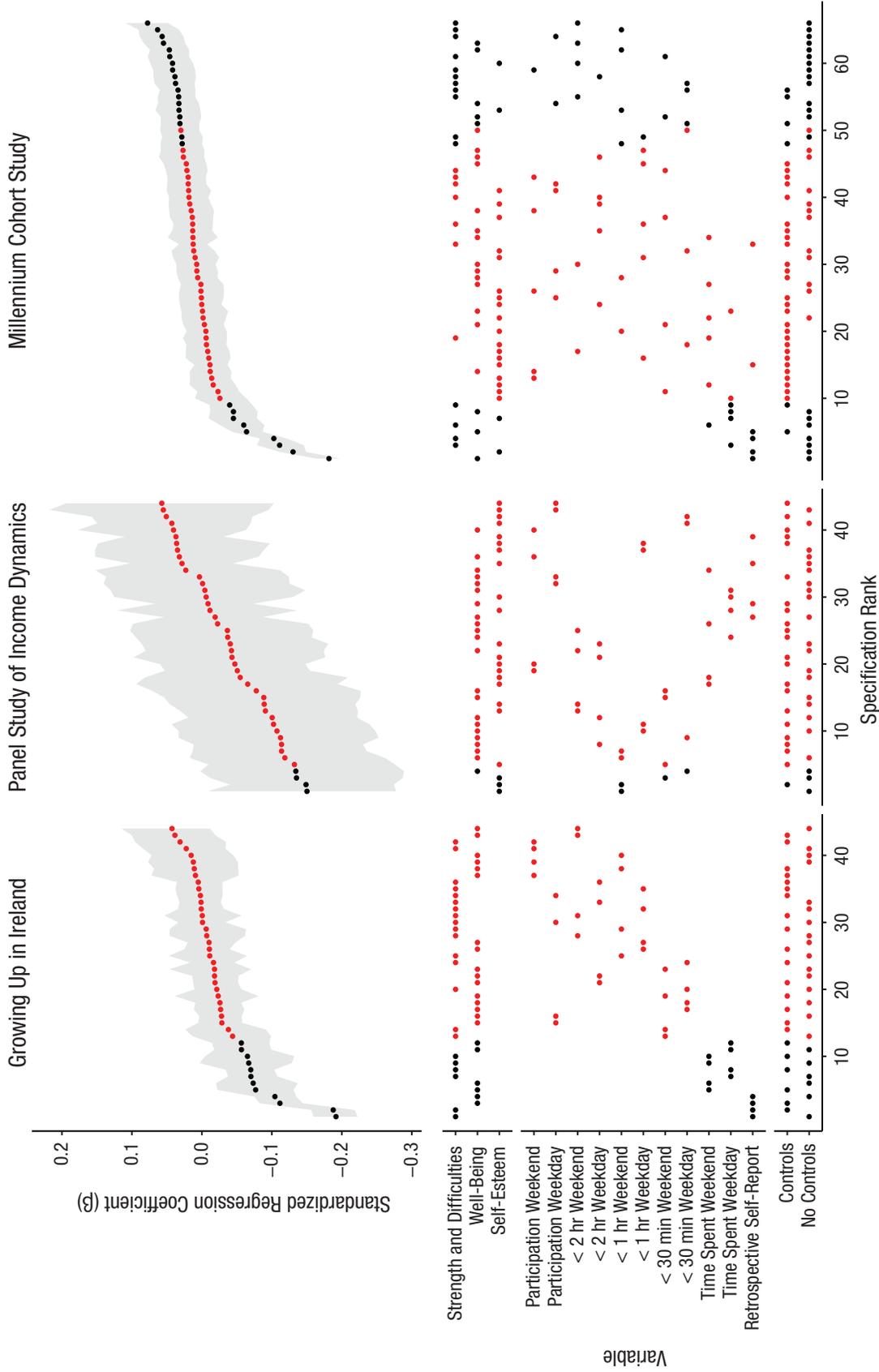


Fig. 1. Results of the specification-curve analysis of Irish, U.S., and U.K. data sets for the association between well-being and digital-screen engagement, calculated using a linear regression. Each point on the x-axis represents a different combination of analytical decisions noted on the y-axis of the bottom half of the graph (the “dashboard”). The resulting standardized regression coefficient is displayed in the top half of the graph. The gray areas denote 95% confidence intervals of standardized regression coefficients obtained using bootstrapping. Specifications marked in red were not statistically significant ($p < .05$); those marked in black were significant ($p < .05$).

proposed by the authors—as the nature of the data meant that these tests did not give an accurate overview of the data (see the Supplemental Material). It was possible to calculate whether the amount of significant specifications or size of the median point estimates found in the original data set was surprising—that is, whether less than 5% of the null-hypothesis data sets had more significant specifications in the dominant direction, or more extreme median point estimates, than the original data set.

To create the data sets in which the null hypothesis was true, we extracted the regression coefficient of interest (b), multiplied it by the technology-use measure, and subtracted it from the well-being measure. We then used these values as our dependent well-being variable in a data set in which we now know the effect of interest not to be present. We then ran 500 bootstrapped SCAs using this data. As the bootstrapping operation was repeated 500 times, it was possible to examine whether each bootstrapped data set (in which the null hypothesis was known to be true) had more significant specifications or more extreme median point estimates than the original data set. To obtain the p value of the bootstrapping test, we divided the number of bootstraps with more significant specifications in the dominant direction or more extreme median point estimates than the original data set by the overall number of bootstraps.

Results

Correlations between retrospective reports and time-use-diary estimates. For Irish adolescents, the correlation of measures relating to digital engagement, operationalized using the time-use-diary estimate (prior to dichotomization into participation and time spent) and retrospective self-report measurement, was small ($r = .18$). For American adolescents, the correlations relating time-use-diary measures on a weekday and weekend day to self-report digital-engagement measurements were small as well ($r = .08$ and $r = .05$, respectively).

Identifying and implementing specifications. We identified 44 specifications each for the Irish and U.S. data sets. For details about these specifications, see the columns in Table 1 for these two data sets. After all analytical pathways specified in the previous step were implemented, it was evident that there were significant specifications present in both data sets (Fig. 1, left and middle panels). Some specifications showed significant negative associations ($k = 16$), though there was a larger proportion of nonsignificant specifications present ($k = 72$). No statistically significant specifications were positive. Specifications using retrospective self-report digital-engagement measures resulted in the largest negative associations in the Irish data. We did not find this trend

in the U.S. data, possibly because of the restricted range of response anchors connected to their self-report digital-engagement measures.

Statistical inferences. Using bootstrapped null models, we found significant correlations between digital engagement and psychological well-being in both the Irish and American data sets (Table 2). We count those correlations as significant that showed significant effects both for the median point estimates and the number of significant tests in the dominant direction.

There was a significant correlation between retrospective self-report digital engagement (median $\beta = -0.15$, $p < .001$; number of significant results in dominant direction = 4/4, $p < .001$) and adolescent well-being in the Irish data set. There were also negative associations for some of the time-use-diary measures, notably time spent using digital screens on a weekend (median $\beta = -0.07$, $p < .001$; number of significant results = 4/4, $p < .001$) and on a weekday (median $\beta = -0.06$, $p < .001$; number of significant results = 4/4, $p < .001$). In the American data set we found significant associations only for digital engagement 1 hr before bedtime on a weekend day (median $\beta = -0.13$, $p < .001$; number of significant results = 2/4, $p = .010$). There were no significant associations of retrospective self-reported digital engagement. Taking this pattern of results as a whole, we derived a series of promising data- and theory-driven hypotheses to test in a confirmatory study.

Confirmatory Study

From the two exploratory studies detailed above, and from previous literature about the negative effects of technology use during the week on well-being (Harbard et al., 2016; Levenson et al., 2017; Owens, 2014), we derived five specific hypotheses concerning digital engagement and psychological well-being. Our aim was to evaluate the robustness of these hypotheses in a third representative adolescent cohort. To this end, we pre-registered our data-analysis plan using the Open Science Framework (<https://osf.io/wrh4x/>), focusing on the data collected as part of the Millennium Cohort Study (MCS; University of London, Institute of Education, 2017), prior to the date that the data were made available to researchers. We made five hypotheses:

Hypothesis 1: Higher retrospective reports of digital engagement would correlate with lower observed levels of adolescent well-being.

Hypothesis 2: Total time spent engaging with a digital screen, derived from time-use-diary measures, would correlate with lower observed levels of adolescent well-being.

Table 2. Results of the Specification-Curve Analysis Bootstrapping Tests for the Irish, U.S., and U.K. Data Sets

Technology measure	Ireland				United States				United Kingdom				Aggregate	
	Median point estimate		Share of significant results in predominant direction		Median point estimate		Share of significant results in predominant direction		Median point estimate		Share of significant results in predominant direction		Median point estimate (β)	Share of significant results in predominant direction
	β	p	Number	p	β	p	Number	p	β	p	Number	p		
Participation: weekend	0.02 [-0.02, 0.05]	.31	0	1.00	-0.01 [-0.07, 0.05]	.47	0	1.00	0.01 [-0.01, 0.02]	.28	1	.32	0.01	1
Participation: weekday	-0.01 [-0.03, 0.01]	.16	0	1.00	0.03 [-0.05, 0.10]	.32	0	1.00	0.02* [0.00, 0.04]	.01*	2*	.02*	0.01	2
Less than 2 hr: weekend	0.02 [-0.02, 0.02]	.27	0	1.00	-0.07 [-0.11, 0.05]	.04	0	1.00	0.04* [0.00, 0.03]	.00*	4*	.00*	0.00	4
Less than 2 hr: weekday	-0.01 [-0.02, 0.01]	.39	0	1.00	-0.07 [-0.14, 0.00]	.08	0	1.00	0.02 [0.00, 0.03]	.01	1	.27	0.00	1
Less than 1 hr: weekend	0.00 [-0.01, 0.05]	.91	0	1.00	-0.13* [-0.15, 0.00]	.00*	2*	.01*	0.03* [0.02, 0.05]	.00*	4*	.00*	0.00	4
Less than 1 hr: weekday	0.00 [-0.04, 0.00]	.65	0	1.00	-0.03 [-0.12, 0.04]	.77	0	1.00	0.02 [0.01, 0.04]	.02	1	.27	0.01	1
Less than 30 min: weekend	-0.03 [-0.06, 0.00]	.05	0	1.00	-0.11 [-0.18, -0.03]	.00	1	.30	0.02* [0.00, 0.03]	.01*	2*	.01*	-0.02	2
Less than 30 min: weekday	-0.02 [-0.03, 0.04]	.01	0	1.00	-0.03 [-0.20, -0.06]	.81	1	.20	0.03* [0.01, 0.05]	.00*	3*	.00*	0.00	3
Time spent: weekend	-0.07* [-0.10, -0.04]	.00*	4*	.00*	-0.04 [-0.11, 0.05]	.24	0	1.00	0.00 [-0.02, 0.01]	.66	1	.25	-0.04	5
Time spent: weekday	-0.06* [-0.08, -0.04]	.00*	4*	.00*	-0.01 [-0.09, 0.05]	.74	0	1.00	-0.04* [-0.06, -0.02]	.00*	4*	.00*	-0.04	8
Self-report	-0.15* [-0.17, -0.13]	.00*	4*	.00*	0.01 [-0.03, 0.05]	.40	0	1.00	-0.08* [-0.10, -0.07]	.00*	4*	.00*	-0.08	8

Note: Values in brackets are 95% confidence intervals. Asterisks indicate results in which both measures of significance are less than .05.

Hypothesis 3: Digital engagement 30 min before bedtime on weekdays, derived from time-use-diary measures, would correlate with lower observed levels of adolescent well-being.

Hypothesis 4: Digital engagement 1 hr before bedtime on weekdays, derived from time-use-diary measures, would correlate with lower observed levels of adolescent well-being.

Hypothesis 5: In models without controls (detailed below), the negative association would be more pronounced (i.e., will have a larger absolute value) than in models with controls.

Method

Data sets and participants. The focus of the confirmatory analyses was the longitudinal MCS, which followed a U.K. cohort of young people born between September 2000 and January 2001 (University of London, Institute of Education, 2017). The survey of interest was administered in 2015 and, after data exclusions, included responses by 11,884 adolescents and their caregivers. It encompassed 5,931 girls and 5,953 boys: 2,864 thirteen-year-olds, 8,860 fourteen-year-olds, and 160 fifteen-year-olds. Using clustered stratified sampling, it oversampled minorities and participants living in disadvantaged areas. Each adolescent completed two (1 weekend day and 1 weekday) paper, Web-based, or app-based time-use diaries within 10 days of the main interviewer visit.

Ethical review. The U.K. National Health Service (NHS) and the London, Northern, Yorkshire, and South-West Research Ethics Committees gave ethical approval for data collection.

Measures.

Adolescent well-being. In addition to the SDQ completed by the caretaker, the Rosenberg Self-Esteem Scale was used, as was an abbreviated version of the short-form Mood and Feelings Questionnaire (Angold, Costello, Messer, & Pickles, 1995). This measure instructed participants as follows: “For each question please select the answer which reflects how you have been feeling or acting in the past two weeks.” Responses were “I felt miserable or unhappy,” “I didn’t enjoy anything at all,” “I felt so tired I just sat around and did nothing,” “I was very restless,” “I felt I was no good any more,” “I cried a lot,” “I found it hard to think properly or concentrate,” “I hated myself,” “I was a bad person,” “I felt lonely,” “I thought nobody really loved me,” “I thought I could never be as good as other kids,” and “I did everything wrong” (1 = *not true*, 2 = *sometimes*, 3 = *true*; scale subsequently reverse scored).

Adolescent technology use. Like the U.S. and Irish data sets, the U.K. data set included a retrospective self-report of digital-screen-engagement items. The mean was taken of the four questions concerned with hours per weekday the adolescent spent in such activities—“watching television programmes or films,” “playing electronic games on a computer or games systems,” “using the internet” at home, and “on social networking or messaging sites or Apps on the internet” (scale ranging from 1 = *none* to 8 = *7 hours or more*).

Using the time-use-diary data, we derived digital-engagement measures in line with the approach used for the exploratory data sets. Participants could select certain activity codes for each 10-min time slot (except in the app-based time diary, which used 1-min slots). Five of these activity codes were used in the aggregate measure of digital engagement: “answering emails, instant messaging, texting,” “browsing and updating social networking sites,” “general internet browsing, programming,” “playing electronic games and Apps,” and “watching TV, DVDs, downloaded videos.”

Covariate and confounding variables. The covariates detailed in the analysis plan were chosen using previous studies as a template (Orben & Przybylski, 2019; Parkes et al., 2013). In both these studies and in the present analysis, a range of sociodemographic factors and maternal characteristics, including the child’s sex and age and the mother’s education, ethnicity, psychological distress (K6 Kessler Scale), and employment, were included as covariates. These factors also included household-level variables such as household income, number of siblings present, whether the father was present, the adolescent’s closeness to parents, and the time the primary caretaker could spend with the children. Finally, there were adolescent-level variables that included reports of long-term illness and negative attitudes toward school. To control for the caretaker’s current cognitive ability, we also included the primary caretaker’s score on a word-activity task, in which he or she was presented with a list of target words and asked to choose synonyms from a corresponding list. While we preregistered the use of the adolescents’ scores on the word-activity task as a control variable, we did not include these because they have previously been linked directly to adolescent well-being or technology use. Furthermore, we included the adolescents’ sex and age as controls in our models; their omission was a clear oversight in our preregistration, in light of the extant literature (e.g., Przybylski & Weinstein, 2017).

Analytic approach

In broad strokes, the confirmatory analytical pathway followed the approach used to examine the exploratory data

sets. We included bootstrapped models of all variables, as in our exploratory analyses, but also extended these to examine the specific preregistered hypotheses. We adapted our preregistered analysis plan to run simple regressions instead of structural equation models to allow us to implement our significance-testing analyses and bootstrapping approaches. Furthermore, after submitting our preregistration, we decided to analyze bootstrapped SCAs instead of permutation tests to obtain CIs and to run two-sided hypothesis tests, rather than one-sided tests, as they are more informative for the reader. We also note that the preregistered data-cleaning code was altered slightly because of previously overlooked coding errors.

In the preregistration, we also specified a smallest effect size of interest (SESOI; Lakens, Scheel, & Isager, 2017), a concept proposed to avoid the problematic overinterpretation of significant but minimal associations, which are becoming increasingly common in large-scale studies of technology-use outcomes (Ferguson, 2009; Orben & Przybylski, 2019). Following Ferguson, we preregistered a correlation coefficient SESOI (r) of .10 (95% CI = [.099, .101]). In other words, digital-engagement associations that explained less than 1% (i.e., $r^2 < .01$) of well-being outcomes were judged, a priori, as being too modest in practical terms to be worthy of extended scientific discussion.

Results

Correlations between retrospective reports and time-use-diary estimates. The correlation between self-report digital engagement and time-use-diary measures of digital engagement was in line with the Irish data ($r = .18$ for both weekdays and weekend days). This was higher and more consistent than what was observed in the U.S. data, as the retrospective self-report response options in the British and Irish data were of better quality.

Identifying and implementing specifications. We identified 66 specifications for the U.K. data set (22 more than for the Irish or the U.S. data) because there were three different measures of psychological well-being and 11 digital-engagement measures; also involved was the decision to include controls or not. For more details regarding these specifications, see the far-right column of Table 1.

The specification results are plotted in the rightmost column of Figure 1. In contrast to the exploratory analyses, this analysis showed significant positive and negative associations between digital engagement and well-being. As with the exploratory data, retrospective self-report measures consistently showed the most negative correlations. Digital-engagement-before-bedtime

measures showed significant positive associations ($k = 18$) and no significant negative associations.

Statistical inferences. Using a similar analytical approach to that used with the exploratory data sets (see Table 2), we found that there was a significant negative correlation between retrospective self-report digital engagement and adolescent well-being (median $\beta = -0.08$, $p < .001$; significant results = 4/6, $p < .001$). There was also a negative association of time spent engaging with digital screens on a weekday (median $\beta = -0.04$, $p < .001$; significant results = 4/6, $p < .001$). There were, however, significant positive associations for other time-use-diary measures of digital engagement, including participation with digital screens on a weekday (median $\beta = 0.02$, $p = .010$; significant results = 2/6, $p = .020$), digital engagement 30 min before bedtime on a weekday (median $\beta = 0.03$, $p < .001$; significant results = 3/6, $p < .001$) and weekend day (median $\beta = 0.02$, $p = .010$; significant results = 2/6, $p = .010$), digital engagement 1 hr before bedtime on a weekend (median $\beta = 0.03$, $p < .001$; significant results = 4/6, $p < .001$), and digital engagement 2 hr before bedtime on a weekend (median $\beta = 0.04$, $p < .001$; significant results = 4/6, $p < .001$).

Hypothesis 1: retrospective self-reported digital engagement and psychological well-being. Because we found a significant negative association between retrospective self-report digital engagement and adolescent well-being (median $\beta = -0.08$, 95% CI = [-0.10, -0.07], $p < .001$; significant results = 4/6, $p < .001$), using a two-sided bootstrapped test, our first hypothesis is supported. The median partial r^2 value (.008, 95% CI = [.006, .011]) was below our SESOI (i.e., $r = -.10$) detailed in the preregistered analysis plan. It must, however, be noted that the 95% CI extends above the SESOI.

Hypothesis 2: time spent engaging with digital screens and psychological well-being. We examined general time spent engaging with digital screens, both on a weekend and weekday, using time-use-diary measures and one-sided bootstrapped tests. A significant negative association (median $\beta = -0.02$, 95% CI = [-0.04, -0.01], $p < .001$; significant results = 5/12, $p < .001$) was in evidence. The direction and significance of this correlation was in line with the registered hypothesis, yet this association was smaller than the prespecified SESOI (partial $r^2 = .001$, 95% CI = [.000, .002]). Again, the 95% CI of the effect size fell above the SESOI.

Hypothesis 3: technology use 30 min before bedtime on a weekday and psychological well-being. Using a two-sided bootstrapped test, we found that results focusing on digital engagement 30 min before bedtime indicated

that Hypothesis 3 was not confirmed (median $\beta = 0.03$, 95% CI = [0.01, 0.05], $p < .001$; significant results = 3/6, $p < .001$), as the effect was in the opposite direction.

Hypothesis 4: technology use 1 hr before bedtime on a weekday and psychological well-being. Models examining the association of digital engagement 1 hr before bedtime on a weekday found no effect in the hypothesized negative direction (median $\beta = 0.02$, 95% CI = [0.01, 0.04], $p = 0.02$; significant results = 1/6, $p = .27$), so the fourth hypothesis was not supported.

Hypothesis 5: comparing models that do and do not account for confounding variables when testing the relation between digital-screen engagement and psychological well-being. Lastly, the effect of including controls in our confirmatory models was evaluated. Figure 2 presents two different specification curves, one including and one excluding controls. Visual inspection of the models and Table 3 shows that the models with controls exhibited less extreme negative associations. This was supported by the difference in the median associations (controls: $r = .026$, no controls: $r = .001$). A one-sided paired-samples t test comparing the correlation coefficients found using specifications with controls to those found with no controls indicated a nonsignificant association, $t(32) = 0.26$, $p = .79$. This result does not support the hypothesis that correlations present when controls are not included in the model are more negative than when controls are included. As we see in Table 4, the five most extreme negative specifications become less negative if controls are added, but the five most extreme positive specifications become less positive if controls are added.

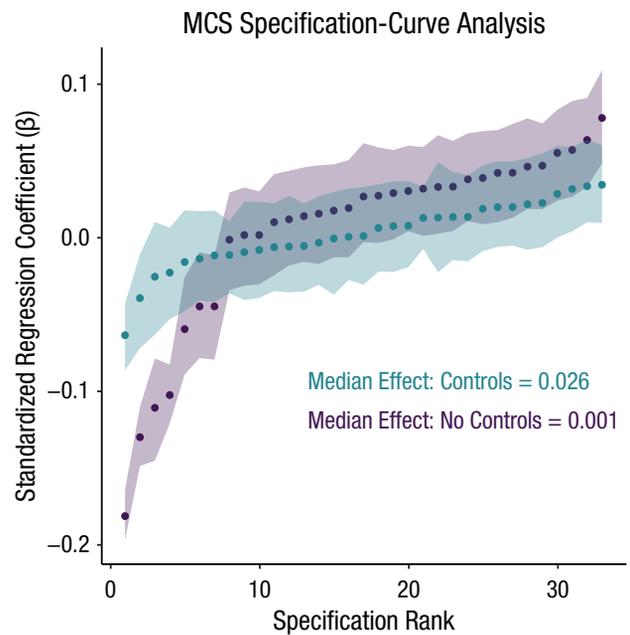


Fig. 2. Results of the specification-curve analysis for the Millennium Cohort Study (MCS) data set. The plot shows the standardized regression coefficients of the linear regressions linking digital engagement and adolescent well-being. The two different curves represent specifications with (teal) and without (purple) control variables. The shaded areas indicate the 95% confidence intervals calculated using bootstrapping.

Discussion

Because technologies are embedded in our social and professional lives, research concerning digital-screen

Table 3. Results of the Specification-Curve-Analysis Bootstrapping Tests for Confirmatory Tests

Technology measure	Median point estimate			Share of significant results in predominant direction	
	β	Partial r^2	p	Number	p
Self-report	-0.08 [-0.10, -0.07]	.008 [.006, .011]	.00	4	.00
Time spent	-0.02 [-0.04, -0.01]	.001 [.000, .002]	.00	5	.00
Less than 30 min on weekday	0.03 [0.01, 0.05]	.001 [.000, .003]	.00	3	.00
Less than 1 hr on weekday	0.02 [0.01, 0.04]	.001 [.000, .003]	.02	1	.27

Note: Values in brackets are 95% confidence intervals.

Table 4. Summary of the Five Most Negative and Five Most Positive Specifications With and Without the Inclusion of Control Variables

Outcome and predictor	No controls		Controls		Difference in β
	β	r^2	β	r^2	
Strengths and Difficulties Questionnaire					
< 1 hr weekend	0.063	.005	0.028	.001	76.8%
< 2 hr weekend	0.078	.008	0.033	.002	80.2%
Participation weekday	0.057	.004	0.020	.001	97.5%
Self-reported	-0.103	.011	0.012	.000	-200.0%
Time spent weekday	-0.111	.016	-0.040	.003	-94.7%
Time spent weekend	-0.060	.005	-0.006	.000	-163.4%
Well-being					
< 1 hr weekend	0.047	.002	0.006	.000	154.4%
< 2 hr weekend	0.055	.003	0.007	.000	152.3%
Self-esteem	-0.130	.017	-0.012	.000	-167.4%
Well-being	-0.182	.033	-0.064	.005	-96.0%

use and its effects on adolescent well-being is under increasingly intense scientific, public, and policy scrutiny. It is therefore essential that the psychological evidence contributing to the available literature be of the highest possible standard. There are, however, considerable problems, including measurement issues, lack of transparency, little confirmatory work, and overinterpretation of miniscule effect sizes (Orben & Przybylski, 2019). Only a few studies regarding technology effects have used a preregistered confirmatory framework (Elson & Przybylski, 2017; Przybylski & Weinstein, 2017). No large-scale, cross-national work has tried to move away from retrospective self-report measures to gauge time spent engaged with digital screens, yet it has been evident for years that such self-report measures are inherently problematic (Scharkow, 2016; Schwarz & Oyserman, 2001). Until these three facts are reconciled in the literature, exploratory studies wholly dependent on retrospective accounts will command an outsized share of public attention (Cavanagh, 2017).

This study marks a novel contribution to the psychological study of technology in a variety of ways. First, we introduced a new measurement of screen time, implemented rigorous and transparent approaches to statistical testing, and explicitly separated hypothesis generation from hypothesis testing. Given the practical and reputational stakes for psychological science, we argue that this approach should be the new baseline for researchers wanting to make scientific claims about the effects of digital engagement on human behavior, development, and well-being.

Second, the study found little substantive statistically significant and negative associations between digital-screen engagement and well-being in adolescents. The most negative associations were found when both

self-reported technology use and well-being measures were used, and this could be a result of common method variance or noise found in such large-scale questionnaire data. Where statistically significant, associations were smaller than our preregistered cutoff for a practically significant effect, though it bears mention that the upper bound of some of the 95% CIs equaled or exceeded this threshold. In other words, the point estimate was below the SESOI of a correlation coefficient (r) of .10, but because the CI overlapped with the SESOI, we cannot confidently rule out the possibility that it accounts for about 1% of covariance in the well-being outcomes. This is in line with results from previous research showing that the association between digital-technology use and well-being often falls below or near this threshold (Ferguson, 2009; Orben & Przybylski, 2019; Twenge, Joiner, Rogers, & Martin, 2017; Twenge, Martin, & Campbell, 2018). We argue that these effects are therefore too small to merit substantial scientific discussion (Lakens et al., 2017).

This supports previous research showing that there is a small significant negative association between technology use and well-being, which—when compared with other activities in an adolescent's life—is miniscule (Orben & Przybylski, 2019). Extrapolating from the median effects found in the MCS data set, we point out that those adolescents who reported technology use would need to report 63 hr and 31 min more of technology use a day in their time-use diaries to decrease their well-being by 0.50 standard deviations, a magnitude often seen as a cutoff for effects that participants would be subjectively aware of (Norman, Sloan, & Wyrwich, 2003; calculations included in the Supplemental Material). Whether smaller effects, even when not noticeable, are important is up for debate, as technology use

affects a large majority of the population (Rose, 1992). The above calculation is based on the median of calculated effect sizes, but if we consider only the specification with the maximum effect size, the time an adolescent needs to spend using technology to experience the relevant decline in well-being decreases to 11 hr and 14 min per day.

Third, this study was also one of the first to examine whether digital-screen engagement before bedtime is especially detrimental to adolescent psychological well-being. Public opinion seems to be that using digital screens immediately before bed may be more harmful for teens than screen time spread throughout the day. Our exploratory and confirmatory analyses provided very mixed effects: Some were negative, while others were positive or inconclusive. Our study therefore suggests that technology use before bedtime might not be inherently harmful to psychological well-being, even though this is a well-worn idea both in the media and in public debates.

Limitations

While we aim to implement the best possible analyses of the research questions posed in this article, there are issues intrinsic to the data that must be noted. First, time-use diaries as a method for measuring technology use are not inherently problem free. It is possible that reflexive or brief uses of technology concurrent with other activities are not properly recorded by these methods. Likewise, we cannot ensure that all days that were under analysis were representative. To address both issues, one would need to holistically track technology use across multiple devices over multiple days, though doing this with a population-representative cohort would be extremely resource intensive (Wilcockson, Ellis, & Shaw, 2018). Second, it is important to note that the time-use-diary and well-being measures were not collected on the same occasion. Because the well-being measures inquired about feelings in general, not simply about feelings on the specific day of questioning, the study assumed that the correlation between both measures still holds, reflecting links between exemplar days and general experiences. Finally, it bears mentioning that the study is correlational and that the directionality of effects cannot, and should not, be inferred from the data.

Conclusion

Until they are displaced by a new technological innovation, digital screens will remain a fixture of human experience. Psychological science can be a powerful tool for quantifying the association between screen use and adolescent well-being, yet it routinely fails to

supply the robust, objective, and replicable evidence necessary to support its hypotheses. As the influence of psychological science on policy and public opinion increases, so must our standards of evidence. This article proposes and applies multiple methodological and analytical innovations to set a new standard for quality of psychological research on digital contexts. Granular technology-engagement metrics, large-scale data, use of SCA to generate hypotheses, and preregistration for hypothesis testing should all form the basis of future work. To retain the influence and trust we often take for granted as a psychological research community, robust and transparent research practices will need to become the norm—not the exception.

Action Editor

Brent W. Roberts served as action editor for this article.

Author Contributions

A. Orben conceptualized the study with regular guidance from A. K. Przybylski. A. Orben completed the statistical analyses and drafted the manuscript; A. K. Przybylski gave integral feedback. Both authors approved the final manuscript for publication.

ORCID iDs

Amy Orben  <https://orcid.org/0000-0002-2937-4183>

Andrew K. Przybylski  <https://orcid.org/0000-0001-5547-2185>

Acknowledgments

The Centre for Longitudinal Studies, UCL Institute of Education, collected Millenium Cohort Study data; the UK Data Archive/UK Data Service provided the data. They bear no responsibility for our analysis or interpretation. We thank J. Rohrer for providing the open-access code on which parts of our analyses are based.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The National Institutes of Health (R01-HD069609/R01-AG040213) and the National Science Foundation (SES-1157698/1623684) supported the Panel Study of Income Dynamics. The Department of Children and Youth Affairs funded Growing Up in Ireland, carried out by the Economic and Social Research Institute and Trinity College Dublin. A. Orben was supported by a European Union Horizon 2020 IBSEN Grant; A. K. Przybylski was supported by an Understanding Society Fellowship funded by the Economic and Social Research Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/10.1177/0956797619830329>

Open Practices



The data can be accessed using the following links, which requires the completion of a request or registration form—Growing Up in Ireland: <http://www.ucd.ie/issda/data/guichild/>; Panel Study of Income Dynamics: <https://simba.isr.umich.edu/U/Login.aspx?TabID=1>; Millennium Cohort Study: <https://beta.ukdata.service.ac.uk/datacatalogue/series/series?id=2000031#!/access>. The analysis code for this study has been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/rkb96/>. The design and analysis plans were preregistered at <https://osf.io/wrh4x/>. Changes were made after preregistration because improved methods became available. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797619830329>. This article has received the badges for Open Materials and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

Note

1. The values on the scale were 0 (0 min), 1 (1–30 min), 2 (31–60 min), 3 (61–90 min), 4 (91–120 min), 5 (121–150 min), 6 (151–180 min), 7 (181–210 min), 8 (211–240 min), 9 (241–270 min), 10 (271–300 min), 11 (301–330 min), 12 (331–360 min), 13 (361 or more min).

References

- Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-world smartphone use. *PLOS ONE*, *10*(10), Article e0139004. doi:10.1371/journal.pone.0139004
- Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, *5*, 237–249.
- Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of Internet use. *Communication Methods and Measures*, *11*, 173–190. doi:10.1080/19312458.2017.1317337
- Bell, V., Bishop, D. V. M., & Przybylski, A. K. (2015). The debate over digital technology and young people. *BMJ*, *351*, h3064. doi:10.1136/BMJ.H3064
- Boase, J., & Ling, R. (2013). Measuring mobile phone use: Self-report versus log data. *Journal of Computer-Mediated Communication*, *18*, 508–519. doi:10.1111/jcc4.12021
- Cain, N., & Gradisar, M. (2010). Electronic media use and sleep in school-aged children and adolescents: A review. *Sleep Medicine*, *11*, 735–742. doi:10.1016/j.sleep.2010.02.006
- Cavanagh, S. R. (2017, August 6). No, smartphones are not destroying a generation. *Psychology Today*. Retrieved from <https://www.psychologytoday.com/blog/once-more-feeling/201708/no-smartphones-are-not-destroying-generation>
- David, M. E., Roberts, J. A., & Christenson, B. (2018). Too much of a good thing: Investigating the association between actual smartphone use and individual well-being. *International Journal of Human-Computer Interaction*, *34*, 265–275. doi:10.1080/10447318.2017.1349250
- Elson, M., & Przybylski, A. K. (2017). The science of technology and human behavior: Standards, old and new. *Journal of Media Psychology: Theories, Methods, and Applications*, *29*(1), 1–7. doi:10.1027/1864-1105/a000212
- Etchells, P. J., Gage, S. H., Rutherford, A. D., & Munafò, M. R. (2016). Prospective investigation of video game use in children and subsequent conduct disorder and depression using data from the Avon Longitudinal Study of Parents and Children. *PLOS ONE*, *11*(1), e0147732. doi:10.1371/journal.pone.0147732
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532–538. doi:10.1037/a0015808
- Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry: The Journal of Mental Science*, *177*, 534–539. doi:10.1192/BJP.177.6.534
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*, 561–582. doi:10.3758/APP.72.3.561
- Hammer, B. (2012). *Statistical models for time use data: An application to housework and childcare activities using the Austrian time use surveys from 2008 and 1992*. Retrieved from National Transfer Accounts: <https://www.ntaccounts.org/doc/repository/Statistical%20Models%20for%20Time%20Use%20Data%20-%20An%20Application....pdf>
- Hanson, T. L., Drumheller, K., Mallard, J., McKee, C., & Schlegel, P. (2010). Cell phones, text messaging, and Facebook: Competing time demands of today's college students. *College Teaching*, *59*, 23–30. doi:10.1080/87567555.2010.489078
- Harbard, E., Allen, N. B., Trinder, J., & Bei, B. (2016). What's keeping teenagers up? Prebedtime behaviors and actigraphy-assessed sleep over school and vacation. *Journal of Adolescent Health*, *58*, 426–432. doi:10.1016/j.jadohealth.2015.12.011
- Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, *29*, 626–631. doi:10.1016/J.CHB.2012.11.007
- Lakens, D., Scheel, A. M., & Isager, P. M. (2017). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*, 259–269. doi:10.1177/2515245918770963

- Levenson, J. C., Shensa, A., Sidani, J. E., Colditz, J. B., & Primack, B. A. (2017). Social media use before bed and sleep disturbance among young adults in the United States: A nationally representative study. *Sleep, 40*(9). doi:10.1093/sleep/zsx113
- Masur, P. K. (2018). *Situational privacy and self-disclosure: Communication processes in online environments*. Cham, Switzerland: Springer International Publishing.
- Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7*, 221–237. doi:10.1177/1745691612441215
- Neuman, S. B. (1988). The displacement effect: Assessing the relation between television viewing and reading performance. *Reading Research Quarterly, 23*, 414–440. doi:10.2307/747641
- Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life. *Medical Care, 41*, 582–592. doi:10.1097/01.MLR.0000062554.74615.4C
- Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour, 3*, 173–182. doi:10.1038/s41562-018-0506-1
- Orzech, K. M., Grandner, M. A., Roane, B. M., & Carskadon, M. A. (2016). Digital media use in the 2 h before bedtime is associated with sleep variables in university students. *Computers in Human Behavior, 55*, 43–50. doi:10.1016/J.CHB.2015.08.049
- Owens, J. (2014). Insufficient sleep in adolescents and young adults: An update on causes and consequences. *Pediatrics, 134*, e921–e932. doi:10.1542/peds.2014-1696
- Parkes, A., Sweeting, H., Wight, D., & Henderson, M. (2013). Do television and electronic games predict children's psychosocial adjustment? Longitudinal research using the UK Millennium Cohort Study. *Archives of Disease in Childhood, 98*, 341–348. doi:10.1136/archdischild-2011-301508
- Przybylski, A. K., & Weinstein, N. (2017). A large-scale test of the Goldilocks hypothesis. *Psychological Science, 28*, 204–215. doi:10.1177/0956797616678438
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*, 151–161.
- Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2017). Probing birth-order effects on narrow traits using specification-curve analysis. *Psychological Science, 28*, 1821–1832. doi:10.1177/0956797617723726
- Rohrer, J. M., & Lucas, R. E. (2018). Only so many hours: Correlations between personality and daily time use in a representative German panel. *Collabra: Psychology, 4*(1). doi.org/10.1525/collabra.112
- Rose, G. (1992). *The strategy of preventative medicine*. Oxford, England: Oxford University Press.
- Scharkow, M. (2016). The accuracy of self-reported Internet use—A validation study using client log data. *Communication Methods and Measures, 10*, 13–27. doi:10.1080/19312458.2015.1118446
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation, 22*, 127–160. doi:10.1177/109821400102200202
- Scott, H., Biello, S. M., & Cleland, H. (2018). Identifying drivers for bedtime social media use despite sleep costs: The adolescent perspective. *PsyArXiv*. doi:10.31234/osf.io/2xb36
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. *SSRN*. doi:10.2139/ssrn.2694998
- Smith, S., Ferguson, C., & Beaver, K. (2018). A longitudinal analysis of shooter games and their relationship with conduct disorder and self-reported delinquency. *International Journal of Law and Psychiatry, 58*, 48–53. doi:10.1016/J.IJLP.2018.02.008
- Survey Research Center, Institute for Social Research, University of Michigan. (2018). Panel Study of Income Dynamics [Data set].
- Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2017). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. *Clinical Psychological Science, 6*, 3–17. doi:10.1177/2167702617723376
- Twenge, J. M., Martin, G. N., & Campbell, W. K. (2018). Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology. *Emotion, 18*, 765–780. doi:10.1037/emo0000403
- UK Commons Select Committee. (2017, April). *Impact of social media and screen-use on young people's health inquiry launched*. Retrieved from <https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2017/impact-of-social-media-young-people-17-19/>
- University of London, Institute of Education, Centre for Longitudinal Studies. (2017). *Millennium Cohort Study: Sixth survey, 2015*. Essex, United Kingdom: UK Data Service. doi:10.5255/UKDA-SN-8156-2
- Vanden Abeele, M., Beullens, K., & Roe, K. (2013). Measuring mobile phone use: Gender, age and real usage level in relation to the accuracy and validity of self-reported mobile phone use. *Mobile Media & Communication, 1*, 213–236. doi:10.1177/2050157913477095
- Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., . . . Kross, E. (2015). Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General, 144*, 480–488. doi:10.1037/xge0000057
- Wilcockson, T. D. W., Ellis, D. A., & Shaw, H. (2018). Determining typical smartphone usage: What data do we need? *Cyberpsychology, Behavior, and Social Networking, 21*, 395–398. doi:10.1089/cyber.2017.0652
- Williams, J. S., Greene, S., Doyle, E., Harris, E., Layte, R., McCoy, S., & McCrory, C. (2009). *Growing up in Ireland national longitudinal study of children: The lives of 9-year-olds*. Dublin, Ireland: Office of the Minister for Children and Youth Affairs.
- Wonneberger, A., & Irazoqui, M. (2017). Explaining response errors of self-reported frequency and duration of TV exposure through individual and contextual factors. *Journalism & Mass Communication Quarterly, 94*, 259–281. doi:10.1177/1077699016629372