

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Reply to: Digital Technology and Adolescent Wellbeing

Amy Orben

Emmanuel College, University of Cambridge

MRC Cognition and Brain Sciences Unit, University of Cambridge

Andrew K Przybylski

Oxford Internet Institute, University of Oxford

Department of Experimental Psychology, University of Oxford

1 In their commentary, Twenge et al.¹ wrongly conclude that the small associations between
2 adolescent well-being and digital technology use presented in Orben and Przybylski² are the
3 product of “analytical decisions that resulted in lower effect sizes”. While some concerns
4 raised were already addressed in the original paper’s supplementary materials, we reject the
5 other concerns by providing additional analyses based on the study data.

6 Twenge et al.¹ rightly identify Specification Curve Analysis³ as a powerful tool for
7 exploratory research. SCA highlights how seemingly inconsequential, though equally valid,
8 decisions taken during secondary data analysis can yield divergent results;⁴ its value is rooted
9 in providing transparency and context. That understood, Twenge et al.’s¹ analyses, and with it
10 the commentary’s references to Kelly et al.⁵, are difficult to interpret because the analysis
11 methods of both the commentary and the paper were not consistently reported, not pre-
12 specified prior to data access, and both presented only a subset of defensible analyses that
13 could have been reported.

14 Twenge et al.¹ suggest the approach taken in Orben and Przybylski² minimised the
15 effects reported and proposed alternative analyses they believed would yield larger estimates.
16 For example, they question our selection of a wide range and combination of outcome
17 variables. The supplement in Orben and Przybylski² (Supplementary Table 6²) demonstrates
18 doing so is prudent because researchers (including the commentators⁶) routinely treat novel
19 combinations of individual survey items as outcome variables instead of using validated
20 aggregate scales presented in the available surveys. Such flexibility extends further to the
21 selection of validated scales; for example, Kelly et al.⁵ reported results from only one of the
22 nine well-being measures present in the dataset. While Twenge et al.¹ conclude that the
23 results of this study are “not a case of ‘cherry picking’”, there is little solid basis for this
24 judgement: no rationale is provided for using only this single operationalisation of well-
25 being.

1 As researchers routinely alternate between using individual items, unique averages,
2 and subsets of validated scales as outcomes in their studies, we therefore included all three of
3 these analysis strategies in our SCA approach. Instead of eliminating methodological
4 flexibility, our SCA therefore made the pronounced influence of selective reporting of a
5 subset of results visible. That understood, and in line with the initial request of the
6 commentary¹, analyses present in the supplementary materials of our original paper report
7 results of an SCA using only aggregate scales and Supplementary Figure 5² demonstrates our
8 considering individual items did not change the conclusions of our main analyses.

9 Similarly, another request made by the commentary was directly addressed in our
10 original paper where we detailed that we did not include the hourly technology use scales
11 present in the American MTF dataset; these questions were only asked in conjunction with a
12 low quality one-item life satisfaction measure. These measures were however included in an
13 alternate SCA which was presented in Supplementary Figure 7².

14 The commentator's argued that the addition of certain analyses would have increased
15 the negative correlations found in Orben and Przybylski². To empirically assess the effects of
16 the proposed modifications we re-ran our SCA (see Methods). The results (Figure 1, Table 1)
17 illustrate that the median association and effect size ($\beta = -0.051$ [-0.072, -0.031], percent
18 variance explained = 0.3% [0.2, 0.6]) are not significantly different from those in our original
19 SCA (median $\beta = -0.032$, percent variance explained = 0.4%). Wearing glasses was still
20 more negatively associated with well-being in adolescents than digital technology use (β_{glasses}
21 = -0.061 vs $\beta_{\text{technology}} = -0.051$). Further, the requested self-harm measure did not produce a
22 significantly more negative association than the complete SCA (median $\beta = -0.054$ [-0.064, -
23 0.046]). In line with our most recent work⁷, females showed a more negative association than
24 males (females: median $\beta = -0.069$ [-0.074, -0.065], males: median $\beta = -0.037$ [-0.041, -

0.032]) and technology use predicted nearly one half of one percent of the variability in the well-being of girls.

Further, by applying non-linear modelling methodology as requested (Table 1), we found that the average association between digital technology use and well-being was slightly positive for low-intensity users (median effect = 0.018, bootstrapped 95% CI [0.001, 0.018]) and slightly negative for high-intensity users (median effect = -0.080, bootstrapped 95% [-0.094, -0.076]). This finding conceptually and statistically replicates a previously uncovered pattern of small non-monotonic relationships linking technology use to well-being outcomes⁸. It is important to note that the strength of this association is probably not practically significant: less than 1% of variance in adolescent well-being is accounted by either these positive or negative trends. With this understood, we strongly caution against approaches like those proposed by the commentary authors such as comparing pairs of extreme values. Picking arbitrary cut-offs is likely to yield spurious findings⁹.

Finally, Twenge et al.¹ assert that percentage of variance explained is not a useful metric for evaluating the size of an effect. Firstly, in Orben and Przybylski² we included a wide variety of effect size measurements, leaving it to the reader to select the measure they wish to interpret. Furthermore, unlike Twenge and colleagues¹, we believe that r^2 is a useful effect size measure in the context of our study as it aims to quantify how influential the predictor of interest is while taking into account other predictors of the outcome that were not measured or included in the analyses. Comparing mean differences, as proposed by Twenge et al.¹, requires assuming there are no influential third factors that could be influencing both the amount of technology used and the well-being of adolescents (e.g. socioeconomic status or parenting practices). For example, if we were to compare adolescents who listened to five hours or more of music a day (instead of focusing on technology), we would find that they have 16% lower well-being than those who listened to only ½ an hour or less – on this basis

we might infer, incorrectly, that music has a negative impact on well-being. Instead, we presented comparison specifications (Extended Data 1) to enable readers to gauge, in more detail, the extent to which technology effects are practically significant. These analyses showed, for example, that wearing glasses had a greater negative association with well-being than using digital technologies. Lastly, we have since publication of our original article expanded our effect size approach. For example, we have recently anchored effects to thresholds of subjective awareness^{10,11}, finding that adolescents might have to use screens for upwards of 11 hours and 14 minutes each day before they would be able to perceive any negative effects.

To conclude, on balance we are optimistic that SCA is a powerful tool to facilitate constructive and evidence-based dialogue. We presented new analyses underscoring its value – we tested the additional specifications requested by the commentary authors and the results did not materially affect the conclusions of our original study. By adopting more transparent approaches to analysing data, and sharing how results are affected by different analytical methods we hope the debate about teens and screens will avoid further ideological entrenchment.

Methods

We chose the UK Millennium Cohort Study for our new SCA presented in this reply as it was the dataset also used by Kelly et al.⁵. The new analyses included all the additional specifications requested by the commentators¹: (1) only validated well-being (sub-)scales, (2) a new self-harm outcome measure (an earlier version of the commentary¹ questioned why this outcome was not included in our initial analyses, but it was removed during peer review), (3) removed control variables that the commentary authors took issue with (seen as potential moderators, we believed they were potential confounders), (4) added another specification

1 which only includes ethnicity as a demographic-only control variable, and (5) examined
2 gender separately. We could not adjust our analyses to add the commentary's request to
3 include the improved measures available in the MTF dataset as suitable measures were not
4 present in the MCS dataset.

5 We analysed the non-linear associations using an interrupted regressions approach.
6 We calculated separate regression models for both low- and high-intensity users —
7 adolescents below and above the median in terms of levels of engagement with digital
8 technology (Table 1).

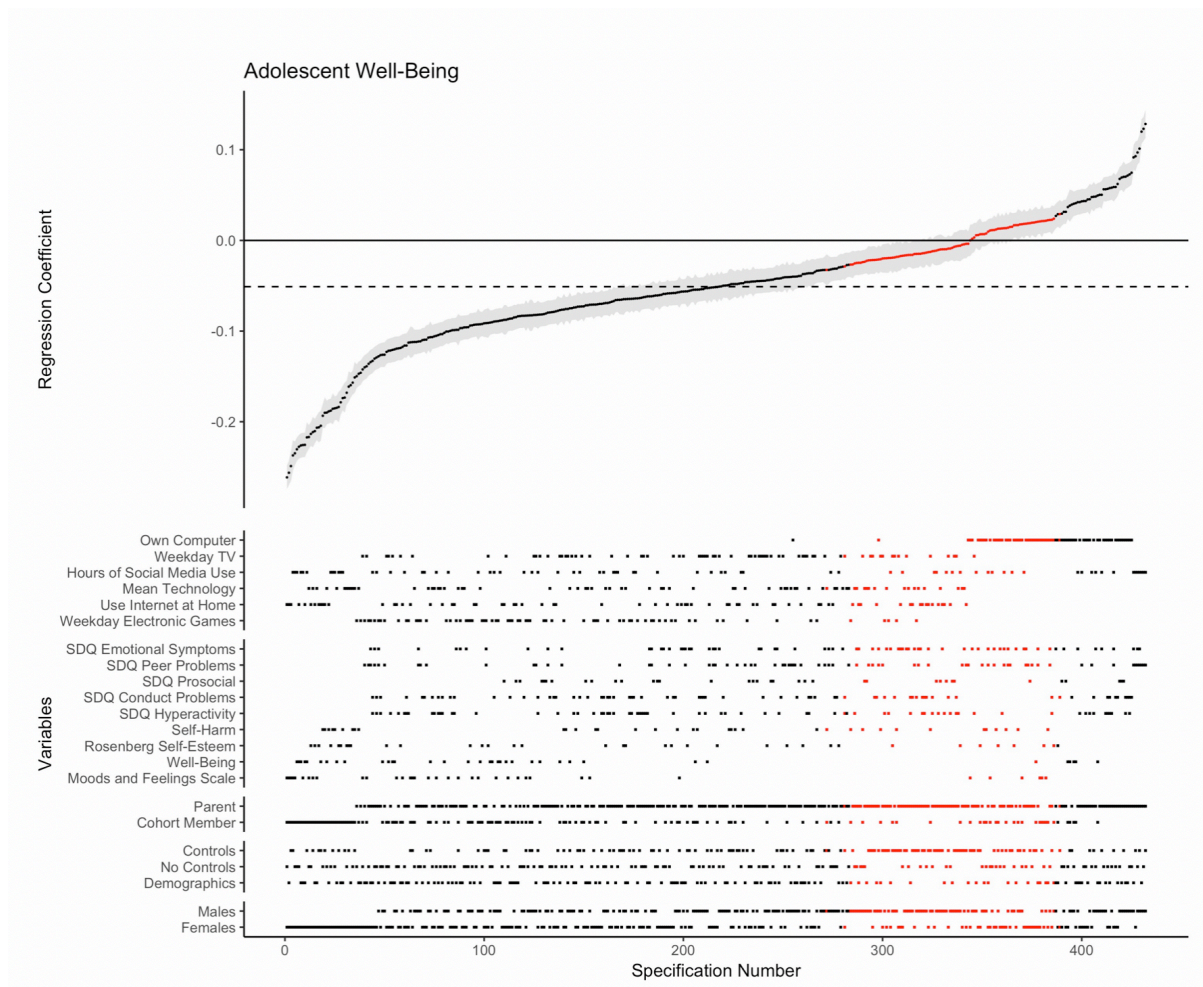
9 The code used to analyse the relevant data can be found on the Open Science
10 Framework (https://osf.io/byqm5/?view_only=f15a442c7225477eb8378c1e717fbd8a). For
11 brevity, we have not included all sample size measurements in this manuscript, but they can
12 be found in the data files on the OSF as well.

13 The data that support the findings of this reply are available from Monitoring the
14 Future (MTF) and the UK data service (MCS) but restrictions apply to the availability of
15 these data, which were used under license for the current study, and so are not publicly
16 available. Data are however available from the relevant third-party repository after agreement
17 to their terms of usage. Information about data collection and questionnaires can be found on
18 the OSF (https://osf.io/7xha2/?view_only=a24cc2ddeceb42bd85dbf3b2babc9a16).
19

1 **Table 1:** Results of the additional SCA analysis implementing the changes suggested by the
2 commentary. The table shows the results both for the complete SCA and specific subgroups.
3 The Confidence Intervals are created by bootstrapping (500 iterations).

Group	Percentage Significant	Median Effect	Median Eta- Squared	Median Number	Median Standard Error
All	75.5 [58.3, 86.8]	-0.051 [-0.072, -0.031]	0.003 [0.002, 0.006]	5272	0.014
Males	69.4 [65.7, 74.1]	-0.037 [-0.041, -0.032]	0.002 [0.002, 0.003]	5248	0.014
Females	81.5 [79.2, 85.0]	-0.069 [-0.074, -0.065]	0.005 [0.005, 0.006]	5295.5	0.014
Controls	58.3 [54.9, 66.0]	-0.033 [-0.039, -0.028]	0.002 [0.002, 0.002]	3491.5	0.016
No Controls	82.6 [79.2, 86.8]	-0.060 [-0.065, -0.054]	0.004 [0.004, 0.005]	5587	0.013
Demographics	85.4 [80.6, 88.2]	-0.058 [-0.065, -0.055]	0.004 [0.004, 0.005]	5272	0.014
Self-Harm	66.7 [61.1, 80.6]	-0.054 [-0.064, -0.046]	0.003 [0.002, 0.004]	5317	0.014
Low Intensity Users	37.6 [31.9, 51.4]	0.018 [0.001, 0.018]	0.001 [0.001, 0.002]	2638	0.020
High Intensity Users	76.4 [72.2, 83.3]	-0.080 [-0.094, -0.076]	0.007 [0.006, 0.009]	2638	0.019

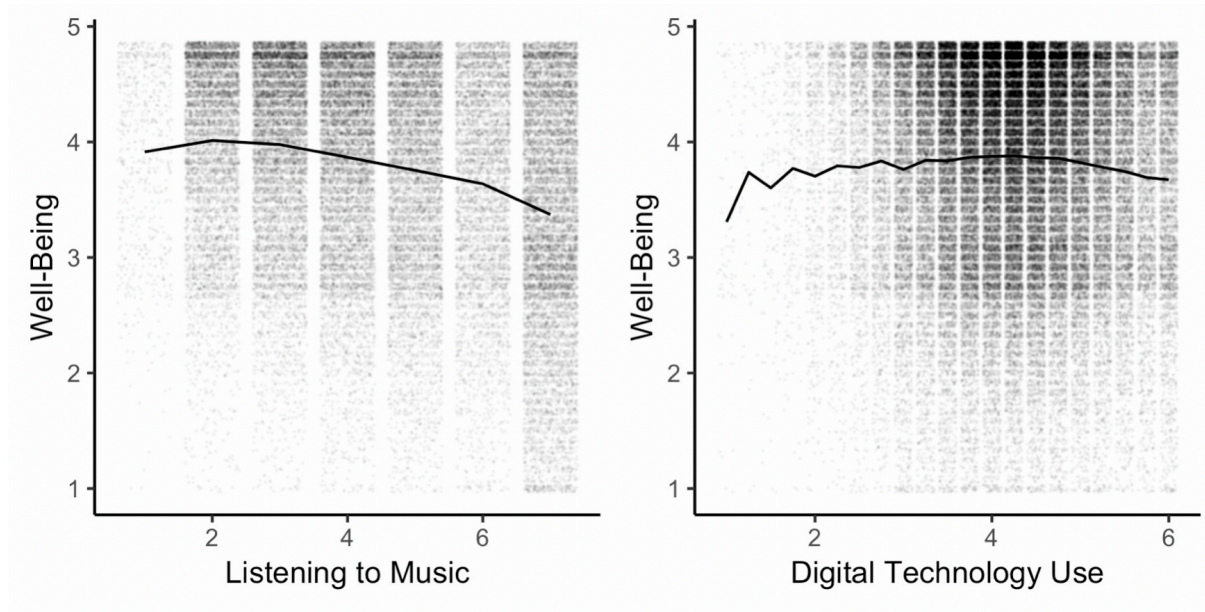
1



2

3 **Figure 2:** Specification Curve Analysis of MCS showing associations between prespecified
4 well-being subscales and technology for males and females. The red squares indicate
5 specifications that were non-significant, while the black squares represent specifications that
6 were significant. The error bars represent the Standard Error. The dotted line indicates the
7 median Regression Coefficient present in the SCA and is included to ease interpretation of
8 the analysis.

9



Extended Data 1: Plots of the variables listening to music and mean digital technology use against the mean of all well-being measures in the survey. The small points are individual participants, while the lines are mean values.

References

1. Twenge, J. M. Reply to: Adolescent technology and well-being. *Nat. Hum. Behav.*
2. Orben, A. & Przybylski, A. K. The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* **3**, 173–182 (2019).
3. Simonsohn, U., Simmons, J. P. & Nelson, L. D. *Specification Curve: Descriptive and inferential statistics on all reasonable specifications*. *SSRN Electronic Journal* (2015) doi:10.2139/ssrn.2694998.
4. Silberzahn, R. *et al.* Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Adv. Methods Pract. Psychol. Sci.* **1**, 337–356 (2018).
5. Kelly, Y., Zilanawala, A., Booker, C. & Sacker, A. Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study. *EClinicalMedicine* **6**, 59–68 (2019).
6. Twenge, J. M., Joiner, T. E., Rogers, M. L. & Martin, G. N. Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time. *Clin. Psychol. Sci.* **6**, 3–17 (2017).
7. Orben, A., Dienlin, T. & Przybylski, A. K. Social media's enduring effect on adolescent life satisfaction. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10226–10228 (2019).
8. Przybylski, A. K. & Weinstein, N. A Large-Scale Test of the Goldilocks Hypothesis. *Psychol. Sci.* **28**, 204–215 (2017).
9. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *BMJ* **332**, 1080 (2006).
10. Orben, A. & Przybylski, A. K. Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. *Psychol. Sci.* **30**, 682–696 (2019).
11. Orben, A. Teens, Screens and Well-Being: An Improved Approach. (2019).

