# A Report on "Digital Distractions with Peer Influence: The Impact of Mobile App Usage on Academic and Labor Market Outcomes" by Barwick et al. (2026)

Reviewer 2

February 07, 2026

v1

# Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made here. Any errors in the report will be corrected in future revisions.

> I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.
>
> Plato, *The Apology of Socrates*, 21d

> To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.
>
> Karl Popper, 'Knowledge and the Shaping of Reality'

# Overview

**Citation:** Barwick, P. J., Chen, S., Fu, C., and Li, T. (2026). Digital Distractions with Peer Influence: The Impact of Mobile App Usage on Academic and Labor Market Outcomes. *Quarterly Journal of Economics*, Vol. 141, Issue 1, pp. 1–49

**URL:** https://doi.org/10.1093/qje/qjaf048

**Abstract Summary:** This paper uses administrative data from a Chinese university to estimate the behavioral and contextual peer effects of mobile app usage on students' academic performance, physical health, and labor market outcomes. The analysis finds that app usage is contagious and harmful across all measured outcomes for both individuals and their peers.

**Key Methodology:** Instrumental Variables (IV) strategy exploiting random roommate assignments, a policy shock (gaming restrictions for minors), and a discrete event (introduction of a blockbuster video game) on administrative student and mobile phone usage data.

**Research Question:** How do students' own and their peers' mobile app usage affect academic performance, physical health, and labor market outcomes, and what are the behavioral and contextual peer effects of app usage?

# Summary

## Is It Credible?

This article by Barwick et al. presents a striking evaluation of the "digital distraction" hypothesis, leveraging a rare combination of administrative university records and high-frequency internet service provider data. The authors argue that mobile app usage—specifically gaming—exerts a substantial causal penalty on human capital formation. Their headline claims are numerically precise and economically alarming: they report that a one standard deviation increase in a student's own app usage reduces their GPA in required courses by 36.2% of a standard deviation and lowers initial post-graduation wages by 2.3% (pp. 2, 6). Furthermore, they contend that digital distraction is contagious, with roommate usage driving down a student's grades through both behavioral spillovers and direct disruption of the study environment (p. 5).

While the direction of these findings aligns with intuition and prior literature, the credibility of the specific effect sizes reported in the headlines is compromised by a methodological oversight in interpretation. The authors estimate models using log-transformed usage hours but calculate the effect of a standard deviation change using a linear approximation that fails to account for the distribution of the data. Because the standard deviation of usage is larger than the mean (p. 14), this approximation significantly inflates the reported magnitude of the effects—likely by more than 50% relative to a precise logarithmic calculation. Consequently, while the negative relationship between gaming and achievement is almost certainly real, the claim that it explains over a third of a standard deviation in GPA appears to be an overstatement derived from this calculation error.

The causal identification strategy is sophisticated but relies on assumptions that warrant scrutiny. The authors utilize the release of the blockbuster game *Yuanshen* and a government policy restricting minor gaming as exogenous shocks (pp. 3–4). For these instrumental variables to be valid, they must affect academic and labor market outcomes *only* through the channel of time spent on mobile apps. This exclusion restriction is debatable. A “blockbuster” game or a national policy discussion could plausibly impact student mental health, social dynamics, or cognitive load directly, independent of the specific minutes logged on a cellular network. Additionally, the article’s effort to disentangle “behavioral” peer effects (contagion) from “contextual” effects (background characteristics) relies on an assumption of homogeneous treatment effects (p. 18). The article’s own supplementary analysis contradicts this, showing that wealthier students react differently to peers (Online Appendix, p. 80), which muddies the interpretation of the “recovered” contextual effects. Finally, the mechanism regarding sleep disruption relies on inferring sleep from periods of phone inactivity (pp. 12–13), a proxy that measures “phone-down time” rather than physiological sleep, potentially conflating rest with other offline activities.

## The Bottom Line

Barwick et al. provide compelling, high-quality evidence that mobile gaming negatively impacts academic performance and early career wages, and that these habits spill over between roommates. However, the headline effect sizes are mathematically inflated, and the true magnitude of the impact is likely considerably smaller than the text suggests. While the causal link is plausible, the precise decomposition of peer effects and the specific policy simulations should be viewed as upper-bound estimates subject to significant uncertainty.

# Potential Issues

**Overstatement of effect sizes for academic outcomes:** The article's interpretation of its main causal findings on GPA appears to be substantially overstated due to the method used to calculate the effect of a one-standard-deviation change in a log-transformed variable. The regression models specify GPA as a function of the logarithm of app usage hours. To interpret the effect, the authors multiply the regression coefficient by the ratio of the standard deviation to the mean of usage hours ($\beta \times SD/Mean$), effectively approximating the change in the log variable with the percentage change in the level variable. Because the standard deviation of monthly app usage (108.5 hours) is larger than the mean (92.9 hours), this approximation is highly inaccurate (p. 14). The more precise calculation for the effect of adding one standard deviation to the mean is $\beta \times \log(1 + SD/Mean)$. In footnote 29 (p. 32), the authors use their method to claim a one-standard-deviation increase in app usage reduces GPA by 0.716 points, equivalent to 36.2% of a standard deviation in GPA. However, the more precise logarithmic calculation yields a reduction of approximately 0.474 points, or 24.0% of a GPA standard deviation. This indicates the reported effect size is inflated by over 50%. A similar overstatement occurs for the OLS estimates (p. 27, footnote 26). While the direction and statistical significance of the effects are not in question, the practical significance of the findings is considerably smaller than claimed.

**Plausibility of the instrumental variable exclusion restriction:** The article's central causal claims rest on two instrumental variables whose validity is debatable, as they may affect outcomes through channels other than the one specified (mobile app usage). First, the release of the game *Yuanshen* is used as a shock that, when interacted with precollege app usage, instruments for in-college usage. The article assumes this shock affects academic outcomes only by changing the time spent on apps (p. 30). However, the introduction of a highly immersive "blockbuster video game" could plausibly have direct effects on students' mental health, social dynamics, or cognitive load, which could in turn affect GPA independently of total usage time. Second, the 2019 minors' game restriction policy is used as a shock that affects students through their underage friends. The article argues that any nationwide effects of the policy, such as those from public discourse, would be absorbed by time fixed effects (pp. 3–4). However, it is possible that the national conversation about gaming addiction had a differential effect on students with more underage friends, perhaps by increasing their awareness or the social stigma of gaming. If so, the instrument would capture this direct psychological effect, violating the exclusion restriction. The authors provide placebo tests showing the instruments do not affect unrelated app categories, but these tests cannot rule out direct effects on unobserved psychological or social factors.

**Methodological issues in the separation of peer effects:** The article claims to separately identify behavioral (contagion) and contextual (peer characteristics) effects, but the procedure used is undermined by the article's own findings. The method involves estimating a reduced-form effect, using an IV to isolate the behavioral component, and then subtracting this component to recover the contextual effect. The authors acknowledge that this procedure relies on a strong assumption of homogeneous peer effects and that their IV estimates a Local Average Treatment Effect (LATE), which may differ from the Average Treatment Effect (ATE) underlying the reduced-form model (p. 18, footnote 20; Online Appendix, p. 57). The article's own heterogeneity analysis then provides evidence against the homogeneity assumption, showing that behavioral spillovers are significantly stronger for students from wealthier families (Online Appendix, p. 80, Table C.15). This suggests that subtracting the estimated LATE from the reduced-form estimate yields a residual that cannot be cleanly interpreted as the contextual effect. Furthermore, the article's definition of the "recovered" contextual effect is non-standard, as it is defined to include "effects of roommates' nonapp usage behaviors" (p. 26, footnote 24), rather than being limited to pre-determined characteristics. This combination of a potentially invalid procedure and an ambiguous definition casts doubt on the conclusion that contextual effects are "much smaller and statistically insignificant" (p. 26).

**Construct validity of key measures:** The article's conclusions depend on proxies for key constructs that may not accurately capture the phenomena of interest. First, sleep patterns are not measured directly but are inferred from periods of low mobile phone activity at night (pp. 12–13). This algorithm defines sleep based on phone non-usage, but a student could put their phone down to study, read, socialize, or be unable to sleep. The measure therefore captures "phone-down time" rather than sleep itself, meaning the conclusion that app usage "reduces sleep duration" (p. 42) may be an overstatement. Second, the primary measure of app usage is derived from a single cellular carrier's records. The article does not specify whether this includes data transmitted over Wi-Fi, which is widely available on university campuses and often preferred for data-intensive activities like gaming. If heavy users systematically shift their activity to unmeasured Wi-Fi networks, the data would suffer from systematic measurement error, potentially understating usage for the most affected individuals and biasing the estimated coefficients.

**Potential for incorrect inference in wage models:** In the analysis of labor market outcomes, the authors include a proxy for unobserved ability derived from the estimated student fixed effect in the GPA regressions (p. 34). This is known as a "generated regressor," as it is an estimate from a first-stage regression that contains its own estimation error. Standard econometric practice requires adjusting the standard errors in the second-stage regression to account for this first-stage uncertainty, for example through bootstrapping. The article does not state that any such correction was performed. Failure to make this adjustment can lead to underestimated standard errors and potentially incorrect statistical inference, casting some doubt on the reported p-values for the wage regressions in Table VI (p. 35).

**Weakness of evidence for sleep as a causal mechanism:** The article proposes that disrupted sleep is a key mechanism through which app usage harms academic outcomes. However, the evidence presented for this mechanism is purely correlational and is disconnected from the main analysis. The authors acknowledge that the sleep data is from a different cohort (2020 only) and a different time period (2023–2024) than the GPA and wage data, and that their instrumental variables predate the sleep data, preventing a causal analysis (p. 42). While the correlations are suggestive, they do not provide direct evidence that sleep disruption was the mechanism responsible for the causal effects on GPA and wages found in the earlier cohorts.

**Limitations to reproducibility and external validity:** The article faces significant limitations regarding both reproducibility and generalizability, which the authors transparently acknowledge. First, the authors did not have direct access to the raw data, instead submitting code to be run by an intermediary in a secure data lab (p. 11, footnote 11). This "remote execution" model, while necessary for privacy protection, makes independent verification of the data processing and analysis impossible, limiting the auditability of the research. Second, the findings are derived from a single "medium-sized, mid-tier" Chinese university where students live in high-density dorms that also serve as primary study spaces (pp. 9–10). The sample also differs from the national student population on key demographics (p. 13). These specific institutional and cultural factors may amplify peer effects and limit the extent to which the findings can be generalized to other settings, such as Western universities or institutions with different living arrangements.

**Minor transparency and presentation issues:** Several minor issues related to transparency and presentation are present in the article. The analysis is based on a sample of students successfully matched to phone records, but 14% of the student population could not be matched, and the article does not provide a comparison of observable characteristics to assess potential selection bias (p. 12). The "back-of-the-envelope" policy calculation is a useful illustration but, as the authors note, it is a partial equilibrium exercise that does not account for how a universal policy might change behavior and social norms (p. 36). The procedure for reweighting the non-representative survey data is mentioned but not fully documented (p. 13). Finally, the authors disclose a procedural issue regarding institutional research ethics, noting that oversight from one of the authors' home institutions was not ceded to the data-holding institution as required by university policy, though the research did receive approval from the local institution's review board (p. 1).

# Future Research

**Direct measurement of physiological mechanisms:** Future work should move beyond inferring sleep patterns from ISP data, which only captures periods of inactivity on a cellular network. Research utilizing wearable technology or actigraphy would provide objective data on sleep duration and quality. This would allow for a rigorous test of the physiological mechanisms proposed in this article, specifically whether digital distractions impair cognitive function through sleep deprivation or through the fragmentation of attention during waking hours.

**Experimental validation of exclusion restrictions:** To address concerns regarding the exclusion restrictions inherent in natural experiment designs (such as game releases), researchers could implement randomized controlled trials involving software-based screen time limits or blocking apps. By directly manipulating the ability to access specific apps without the confounding factors of broad social trends or viral marketing campaigns, such studies could isolate the effect of usage time from the broader psychological impacts of gaming culture.

**On-device data collection:** Future studies should prioritize on-device data collection (e.g., via Screen Time APIs) rather than relying on ISP records. This would address the potential measurement error caused by unobserved Wi-Fi usage, which is common on university campuses. Capturing total screen time across all networks would provide a more accurate denominator for calculating the elasticity of academic performance with respect to digital consumption.

isitcredible.com