Ok, I've read through the Orben paper carefully as well as the original paper that proposes the statistical method they use (SCA). While I admire many of the scientific practices that they use, there are several points that I'd be interested to talk to the authors about. I've got few detailed thoughts on this, but first, so as not to bury the lede, I'll answer your question directly:

What's the evidence that really matters, and did these researchers find it?

If your scientific question has to do with the association between teen mental health and use of smart phones / social media, Orben's paper shows a negative effect on par with bullying. The authors of that study, however, focus on the much broader question of the association between teen well-being and general technology use which ends up diluting the negative effect of social media and showing a substantially weaker effect.

More depth if you're interested:

What question are we interested in?
Orben asks a *very broad* question: "What is the association between the general use of technology (cell phones, computers, TV) on teens' general sense of well-being?" The answer they find is that it's negative, but so small as to be almost ignorable. This isn't surprising to me.

One of the nice thing about their SCA curves, though, is that you can see the answers to narrower questions. If you ask the question, "What is the association between social media use and sense of well-being", you can answer it for the MTF dataset by looking at figure 2 and looking at the leftmost quarter of the curve. The effect size there is -0.04 to -0.05, which is about the same as what they report for binge drinking (-0.045 in the last paragraph of p 176).

Essentially, they're averaging this negative effect against other modes of technology use that I don't think anyone thinks are bad, e.g. watching TV on the weekends. In fact, watching TV on the weekends might well be negatively correlated with social media use, which is one explanation for the relatively positive effect. I very much doubt the authors are deliberately trying to lessen the effect of social media use (though its worth noting that Orben has worked at Facebook in an advisory capacity), but if you're trying to loop this paper into the conversation on the effect of smartphones/social media on teen mental health, you'd be best served by looking at the left part of each of the curves, which shows a substantially larger negative effect.

Orben is using the SCA in a much broader set of models than the previous research they cite

The basic idea is to not consider just a single model that might be cherry picked by the researcher to show a particular extreme effect, but to consider all models within some set. The goal is to reduce the "researcher degrees of freedom". There are a ton of things that go into those degrees of freedom, but off the top of my head, I'd think of:

1. Choice of response variable
2. Choice of explanatory variable(s)
3. Functional form linking 1 and 2 (e.g. the % variance explained thing assumes the function is linear).
4. Tweaks to how we define 1 and 2 (ex1: if one of your explanatory variables is "us state", you might want to also consider a variable that aggregates states into regions, or a variable that drops AK and HI. ex2: if your response is "income", you could also consider "log(income)".
5. Tweaks to which observations we consider (e.g. dropping outliers)
6. Tweaks to (3) (e.g. if your response is a 0 - 1 variable like "do you have cancer", there are two slightly different shaped functions that researchers might use: logit and probit).

The [original SCA paper](#) that they cite as well as [another that they refer readers to](#) pretty much assume that researchers agree on 1, 2, and 3. The focus of SCA is to be sure researchers aren't doing small tweaks to cherry pick in making decisions related to 4 - 6 (and some of 2). SCA definitely does not control for all sources of funny business, but it does control for it in the part of the process (4 - 6) where considering all possible models is a reasonable thing to do; that is, where the decisions truly do seem arbitrary. The logit vs probit functions, for example, are generally selected just based on what the researcher tends to use - habit.

A problem, as I see it, is that Orben uses SCA to average over decisions that researchers would consider *not arbitrary*. In fact, they focus almost exclusively on 1 and 2 which are by far the most consequential decisions - they're the ones that determine the scientific question that you can answer. If you're interested in another video call, I could show how Orben's paper is using SCA using the SCA's own diagram (their figure 1).

I'm realizing something else just now. The previous papers show SCA curves that only cover a single regression coefficient (look at fig 3 [here](#)). Orben's fig 2 curve, for example, is combining the curves for TV weekday, TV weekend, Internet for news, Technology Mean and Social media use. This basically brings us back to the first point: Orben is asking *multiple substantively different questions* in this curve and averaging over the answers.

Orben is using SCA in a framework where the causal structure is much more complex that the research they cite

Those previous two SCA papers share the property that there is a single explanatory variable of interest, and that variable is _exogenous_ (this is a term i usually hear economists use). The models that they consider are:

1. deaths ~ femininity x damages + damages
2. hire ~ blackname x quality + blackname + quality
3. (one example:) intellect ~ birthorder + gender

Where the first can be read, "we seek to explain deaths by a hurricane as a function of its severity (measured by proxy by the $$ damages) and by the interaction between that and the femininity of the name (the theory being that people wouldn't take a storm called "nancy" as serious as they would a storm called "magnus", so they wouldn't evacuate and they'd die).

An exogenous explanatory variable means that it's not casually related to anything else in the model besides the response. Since hurricanes are named randomly, the name has nothing to do with damages and is not caused by deaths. Said another way, the only possible causal arrow there is from femininity to deaths. The same thing can be said of the black names, which were randomly assigned to resumes of the same quality and birthorder, which has nothing to do with gender.

(one example:) self-esteem ~ TV_weekday + mother_has_job + household_income

TV_weekday use might affect self-esteem, but self-esteem might affect TV_weekday. Mom being employed might drive down self-esteem and drive up TV_weekday, so even if there was no causal link between tv_weekday and self-esteem, we'd see a negative correlation between TV_weekday and self-esteem.

I like the previous SCA analyses in part because they focus on truly arbitrary analysis decisions (so I trust the average effect that i'm seeing is the one i'm interested in) and they're working in much simpler causal settings, so I have some confidence that that effect is indeed causal. To their credit, Orben does provide a good paragraph with the disclaimers re: causation, the first paragraph of p. 179.

So what about Twenge's point about the "obscure statistic"?
The discussion we had earlier still stands: Orben is working (I think) with all continuous data, so at first blush, the % variance statistic (which can be computed from the Beta regression coefficient) is a fine thing to use. I've had a look at her code (she did a great job of making her work transparent and reproducible) and she appears to fit the linear model pretty blindly and also says in the end notes of her paper that they didn't check distribution assumptions. So if there are non-linear relationships between these variables (which is conceivable), the % variance statistic actually isn't really telling us what we want to know.

One thing that is kinda cool: since she runs the same analysis on other non-related videos (how much fruit do you eat? how much do you binge drink?) and reports % variance for those, you can at least get something out of comparing the effects to one another. You can show, for example, that the effect (as measured by % variance explained) of social media use is on par with bullying.

Seeing as how Amy Orben is so great about transparent research, I'm inclined to follow suit and engage her over github to ask if an interpretation about the rough equivalent between social media use and bullying is warranted by her analysis. This could happen out in the open over the issues tab.