

Decoding Instagram Use: The Impact of Quality Content Engagement and Harmful Content Reduction on Young Adults' Well-Being

Anja Stevic   1

Yi-Chia Wang  1

David de Segovia Vicente  2

Sunny X Liu  1

Jeffrey T Hancock  1

1. Social Media Lab, Department of Communication, Stanford University, Stanford, CA, USA

2. Department of Communication Sciences, Ghent University, Belgium

Published on 01/13/2026

Research Plan

DOI: [10.71240/lcyc.568727](https://doi.org/10.71240/lcyc.568727)



Abstract

Young adults increasingly rely on social media platforms for connection and engagement with both private circles, including close and distant social ties, and public personas. Instagram, in particular, enables both public and private interactions while allowing users to engage with content from various resources. This three-wave longitudinal study aims to examine how different types of content engagement on Instagram influence the psychological well-being of young people. Through two content interventions, we will explore whether user-centric quality content engagement or harmful content disengagement dynamics—such as following more quality content or reducing exposure to harmful content— affect well-being over time. Our research

enhances understanding of whether and how the quality of social media content can affect the psychological well-being of young adults.

Evaluations

SERVICE	SUMMARY	VERSION	DATE	EVALUATED
<i>Response to reviewers</i>		2	01/13/202 6	

Registration

This Research Plan is the study registration.

Materials

Not applicable; the study has not yet started.

Data

Not applicable; the study has not yet started.

Code

Not applicable; the study has not yet started.

Paper

Decoding Instagram Use: The Impact of Quality Content Engagement and Harmful Content Reduction on Young Adults' Well-Being

By: Anja Stevic, Yi-Chia Wang, David de Segovia Vicente, Sunny X. Liu, and Jeffrey T. Hancock

Introduction

Concerns about the impact of social media on youth have been growing in recent years. Social media plays an important role in the lives of young adults, who actively engage with these platforms to shape their identities, navigate offline experiences, acquire knowledge, and foster social connections. Among various platforms, Instagram is one of the most relevant sources of influence. The developmental phase between ages 18 and 25 is marked by increasing independence, transitions to college, changing jobs, and the formation of new social relationships (Bonnie et al., 2015). Relying on Instagram as a source of information, entertainment, and connection became increasingly high among this demographic (Pew Research, 2024) which is going through a critical period for social and personal development (Wood et al., 2018).

Recent research proposes that the content users engage with on social media, rather than the amount of time spent on these platforms, drives well-being outcomes that range from positive aspects such as social support and hedonic well-being to negative effects like social comparison, anxiety, and depression (Hancock et al., 2022). Yet, in a recent in-depth interview study we found that the content adolescents engage with on Instagram primarily reflects their personal interests, ranging from sports and baking to makeup and life hacks. Additionally, our findings have revealed content's fleeting nature. As one adolescent shared with us "I forgot what I see within five minutes of closing the app". Thus, we must ask: in what ways does content matter? Is it the quantity, type, or quality of content that influences well-being outcomes? Our first question is: How does the content young adults engage with affect their psychological well-being?

In addition to examining content, we aim to explore the heterogeneous effects of social media use. A recent report from the National Academy of Sciences in the U.S. concluded that the impact of social media on well-being varies among adolescents. While some adolescents benefit from social media engagement, experiencing positive effects on their well-being, others face significantly adverse outcomes associated with its use (Beyens et al., 2020). Understanding which individuals are most vulnerable to the negative effects of social media and, more importantly, identifying effective interventions to mitigate these harms is a critical area of research.

Existing research has primarily focused on gender differences and found that girls are more likely than boys to experience upward comparison, body dissatisfaction, and mental health issues (Jung et al., 2022; Maes & Vandenbosch, 2022). However, analyses have shown some dispute pertaining to the actual problems users face on Instagram (Raychoudhury, 2021) with both positive and negative experiences reported. Identifying harmful content is important given that many young users encounter various risks online. Understanding self-identified user-centric harms associated with social media use is essential for effectively mitigating harmful content.

World Economic Forum's Global Coalition for Digital Safety (2023) explains types of harms affecting youth: threats to personal and community safety, harm to health and well-being, hate and discrimination, violation of dignity, invasion of privacy, and deception and manipulation. These categories encompass a range of harms that young people experience online, from exposure to inappropriate content and cyberbullying to privacy breaches and online scams. Identifying the frequency of exposure to these risks enables researchers to tailor interventions that could effectively mitigate potential harms and create safer online environment especially for users that are at higher risk.

Due to an increasing concern about social media and its impact on youth, research on specific platforms and their psychological and health-related effects is urgently needed. The Surgeon General has proposed adding a warning label on social media platforms, stating that social media is associated with mental health harms (Murthy, 2023; Haidt, 2024). Parents across the country are worried about the harms and addictive nature of social media use among adolescents (Douglas et al., 2020). Schools in some states have begun enforcing phone bans (Assembly Committee on Education, 2024).

Despite these concerns, there is limited evidence regarding the impact of content on the associations between social media use and well-being, as well as identifying the individuals most vulnerable to its negative effects. This gap persists largely because most research relies on cross-sectional designs and self-reported data, which do not test potential solutions or interventions. The lack of intervention studies on social media represents a significant research gap, as most existing studies focus on correlations rather than causation, therefore restricting our ability to evaluate the effectiveness of engagement and mitigation strategies). Therefore, we aim to test content interventions in this study as an exploratory approach.

Addressing this gap is crucial, as interventions provide actionable insights that go beyond mere correlation, offering evidence-based recommendations for users, policymakers, educators, and platforms. By incorporating experimental and longitudinal intervention study designs using both self-reported and actual behavioral data, the current study can advance the field, bridging the gap

between theory and practical application. This contribution is essential for developing targeted, scalable, and effective digital well-being strategies that can be implemented in real-world settings.

Social Media Engagement and Well-Being

Well-being is a complex multifaceted concept that has been operationalized in many different ways. Researchers focused on general well-being often operationalize this construct as a self-evaluation of how one generally feels (Diener, 1984; 2015). However, the wide variety of psychological experiences that engaging on social media content can elicit, it is crucial to understand how different types of content relate to more specific facets of well-being, as these effects likely differ (Hancock et al., 2022). While many specific well-being outcomes have been connected to social media use in previous research (see Hancock et al 2022 for a review), some of the most relevant ones in the context of Instagram use are self-esteem (Saiphoo et al., 2020) and quality of social relations (Hall, 2020).

Research addressing the impact that social media has on youth's well-being outcomes is generally divided into two main areas (Weinstein, 2018). The first focuses on the benefits of social media use for well-being (Chan, 2015, 2018), while the second examines its potential risks (Keles et al., 2020). While previous studies have dealt mainly with objective social indicators (e.g., family life or household income), recent research has increasingly focused on the measurement of subjective well-being indicators, such as life satisfaction or quality of life, positive and negative experiences, and physical health (e.g., Chan, 2015; Diener et al., 2015; Vuorre & Przybylski, 2024). Self-perceptions, self-esteem and the perceived quality of the relationships, represent additional relevant subdimensions of well-being in youth's lives. This framework integrates both hedonic and eudaimonic dimensions of well-being. The hedonic perspective emphasizes the presence of positive emotions and the absence of negative affect (Kahneman et al., 1999), while the eudaimonic approach focuses on cognitive appraisals of one's life and personal fulfillment (Ryan & Deci, 2001).

Given that social media platforms like Instagram offer multiple ways of usage, it is important to examine how different patterns of use are associated with varied well-being dimensions. In this research, we operationalize well-being with focus on both positive and negative psychological well-being. We define it as a multifaceted construct encompassing both positive psychological adjustment and the absence of maladaptive functioning (Diener, 1984; Keyes, 2002; Ryff & Singer, 1998). This dimension includes positive indicators, such as self-esteem and life satisfaction (Diener et al., 1999), as well as negative indicators such as negative affect and self-evaluation (Keyes, 2005; Seligman, 2011) that are our key outcomes.

Social media use has been shown to have numerous positive effects on different indicators of well-being, particularly when individuals engage in interactive and targeted activities. One of the key benefits of social media is its role in fostering and maintaining social relationships through

messaging and other interactive features (Burke et al., 2011; Hall & Baym, 2012). Social media platforms provide opportunities for individuals to connect with diverse networks, which has been linked to greater psychological well-being (Ellison et al., 2007; Phua et al., 2017; Chan, 2015, 2018). A commonly used framework in recent research to understand when social media use has positive (rather than negative) effects on well-being is the active/passive dichotomy (Verduyn et al., 2022), where active social media use is generally thought to influence well-being outcomes positively. While evidence in support of this dichotomy is mixed (Valkenburg et al., 2022), it seems clear that many of the positive mechanisms linked to social media use come from active engagement. After all, the affordances of social media platforms, such as targeted communication and one-click interactions, allow users to maintain connections (Burke & Kraut, 2016). For instance, self-expression and direct interpersonal communication with others is associated with increased well-being (Liu et al., 2019). Engaging in private interpersonal interactions and sharing personal content can lead to feelings of social support and belonging, further strengthening emotional well-being. Overall, the ability to engage meaningfully with others through social media enhances social connectedness and psychological well-being (Marciano et al., 2024), underscoring its potential for positive outcomes in young people's daily lives.

In our study, active engagement is operationalized by counts of different types of interactions (e.g., comments, post creation and sharing, sending and receiving private messages) based on previous research on targeted Facebook use for social interactions (Burke & Kraut, 2016). The researchers analyzed data from Facebook users, linking self-reported well-being measures to their actual Facebook activity logs. They found that receiving personalized, composed messages from close friends—referred to as “strong ties”—was associated with improvements in well-being. In contrast, passively viewing broad broadcasts from friends or receiving generic one-click feedback (like “likes”) did not have the same positive effect. This suggests that meaningful, tailored communication from close connections on social media platforms can enhance psychological well-being, while more superficial interactions may not provide the same benefits.

In addition to users' engagement on Instagram, the key factor is the type of content users are exposed to and decide to engage with. Prior research has investigated how different types of content and account engagement relate to wellbeing (Burnell et al., 2023; Depow et al., 2025; O'Kane et al., 2022). For example, Burnell et al. (2023) examined how exposure to either three an acquaintance, an influencer, or one's own profile content relates to psychological well-being, including multiple domains such as positive and negative affect, self-esteem, self-perceptions, and interpersonal negativity, showing that users' engagement with own profile led to positive influence on wellbeing while browsing the profile of either an acquaintance or an Instagram influencer led to negative changes in psychological well-being. In another intervention, Depow et al. (2025) found that users who received brief positive empathy instructions before 10 minutes of browsing their own Instagram feed showed greater affective well-being and life satisfaction relative to participants who were instructed to browse as usual. Both studies imply changes in wellbeing due

to exposure and content based engagement with unique personalized content on Instagram, but do not show how personally relevant and meaningful the content is to users.

Therefore, in this study, we will examine how more engagement with personally identified *quality content* versus less engagement with personally identified *harmful content* on Instagram affects users' well-being over the course of three weeks (short-term longitudinal design). We define quality content and harmful content as content that is personally identified. Quality content represents content that a user identifies as having a positive psychological impact when they are exposed to it and engage with it. Following established characteristics of quality content, we provide a high-level core definition and characteristics that focus on describing the quality content, based on *E-AIMS theoretical framework including* having *meaningful* experiences that relate individuals lives, and *actively interacting* with content in high-quality ways (Hirsch-Pasek et al., 2015). Quality content is engaging by presenting a clear narrative or purpose, actively involving by encouraging meaningful user interaction or reflection, meaningful by resonating with the everyday experiences, values, or identities of the target audience, and socially oriented by fostering dialogue, shared experiences, or collaborative engagement through interactive features such as comments, polls, or live sessions (Hirsh-Pasek et al., 2015). The aim of our study is to encourage active engagement with personalized, self-identified user-centric quality content through identification, following, liking, commenting, and sharing quality content.

Harmful content is defined as content that each user identifies as distressing or negative content in their feed. It is characterized by harmful, toxic, negative, overly idealized, appearance-focused, unrealistic lifestyles that are likely triggering negative self-comparisons. Main characteristics include content that is often highly curated and does not foster meaningful interaction (Alsoubai et al., 2024; Kim et al., 2024).

In the control group, participants will be instructed to use Instagram as usual without any instructions to engage with quality content or harmful content. In the control condition participants are told to browse their feed passively and focus on the content being posted. They will be told not to focus on other sites or apps, but to redirect their attention to the content on their feed if they get distracted. We will randomly assign participants to one of the three conditions. We will measure their wellbeing indicators and exposure to risks on Instagram before the start of the intervention to allow for pre- and post-test. We will continue with the same intervention at Time 2 and again measure the same wellbeing indicators and exposure to risks.

In addition to positive and negative affect as indicators of mood changes, overall life satisfaction is a key dimension of psychological well-being (Diener et al., 1985; Schimmack, 2008). We are interested in exposure to quality content on Instagram bearing in mind divergent effects on psychological well-being depending on the underlying cognitive and emotional processes it activates (Verduyn et al., 2017; Verduyn et al., 2022). On one hand, following and engaging with quality content may foster inspiration, which in turn enhances self-esteem and life satisfaction while reducing negative mood and negative self-evaluation. On the other hand, such exposure

may also trigger social comparison, leading to lower self-esteem (Meier et al., 2020; Valkenburg et al., 2022). Considering the subtle yet significant impact of social media on life satisfaction (Orben & Przybylski, 2019), we hypothesize that participants in the user-centric quality content condition will report higher life satisfaction than those in the control condition. Therefore, we predict:

H1: Participants in the quality content intervention will show improvement in psychological wellbeing.

H1a: Participants in the quality content intervention will show improvement in life satisfaction at week 2 and week 3 compared to the control group, controlling for baseline measure of life satisfaction, and time spent on Instagram (session duration and frequency of social interaction).

H1b: Participants in the quality content intervention will show improvement in mood at week 2 and week 3 compared to the control group, controlling for baseline measure of positive mood, and time spent on Instagram (session duration and frequency of social interaction).

H1c: Participants in the quality content intervention will show lower negative self-evaluation and higher positive self-evaluation at week 2 and week 3 compared to the control group, controlling for baseline measure of negative self-evaluation, depression, self-esteem, loneliness, and time spent on Instagram (session duration and frequency of social interaction).

Exposure to harmful content on Instagram might be detrimental for young users. Harmful online content, including material that promotes self-harm, disordered eating, hate speech, fake news, scams, and unwanted sexual content poses a significant risk to young people's well-being (Ofcom, 2020). Exposure to harmful content has been linked to lower subjective well-being, potentially exacerbating feelings of distress and vulnerability (Keipi et al., 2017). Given the widespread nature of harmful content online, further research is needed to understand its long-term psychological impact and to develop effective interventions for mitigating its effects.

In terms of harmful contact experienced online, studies have shown that online bullying is prevalent in the US especially among marginalized youth (Kreski et al., 2022) resulting in risky offline behaviors. Experiencing online risks has been consistently linked to lower psychological well-being, particularly in terms of self-esteem (Palermi et al., 2017; Patchin & Hinduja, 2010). Many social media users receive cyberbullying messages on Instagram that can harm their well-being (Kim et al., 2014). Longitudinal evidence on this association remains limited, yet findings from a large-scale study of 3,127 children in the United States indicate that cybervictimization can contribute to decreased self-esteem over time (Smokowski et al., 2014). These findings highlight the need for further research into the long-term psychological as well as physical consequences.

It is relevant to take into account self-identified perceived risk experiences of users themselves when designing new interventions. After all, young adult users are the ones using the platforms, with distinct levels of exposure to risks. In the seminal study on Instagram use, Razi et al. (2022) found that 14% of users experience high risk in conversations through flagging conversations with

risks. Due to risk exposure, we expect users will be more receptive to interventions that aim to mitigate the harms they encounter. Agha et al. (2023) study among teens found they often encountered information breaches and sexual risks with strangers, as well as cyberbullying. The usual strategies were blocking or reporting those profiles, however, they did not find any of the existing ways for responding to these risks to be effective in keeping them safe. Researchers have therefore developed intervention ideas co-designed with users teaching teens user experience prototyping tools so they could sketch app features for online risk prevention. Participants used the tools to outline specific online risk experiences and designed features to manage the content they view and other users they could safely interact with. Some teens also designed personalized nudges to handle online risks themselves without getting their parents involved fostering their autonomy (Agha et al., 2023). Similarly, in another study young Instagram users reported using backpedaling strategies to regain control over their feed (Landesman et al., 2024).

Research studies looking at risks and harassment on social media suggest negative outcomes after experiencing harassment through a single direct message (Kim et al., 2024). However, the post-harassment effect washes away after two-weeks demonstrating that the time aspect is relevant to take into account. One of the limitations of previous studies is not looking at cumulative effects of harassments in terms of identifying harms and strategies to mitigate them and reduce exposure to harmful content. In this study, we aim to address the lack of time component by testing the effects of harmful content reduction over time. This way we can understand the duration of the intervention effects better.

Research further suggests the need for targeted risk prevention strategies for youth online (Alsoubai et al., 2024). To contribute to the research gap, in our study we plan to guide Instagram users in identifying harmful content in their own feed. For those users who identify exposure to different types of risks on Instagram in the first survey wave, we are interested in understanding how effective the risk mitigation intervention would be for them over time. Previous research has identified five risk profiles emerging among Instagram users (Alsoubai et al., 2024). The majority of 51% fall into the low risk category, 29% fall into medium risks category, and the other users experienced increased sexting (8%), increased self-harm (8%), and high risk perpetration (4%). We expect a similar outcome in our study, with approximately 20% of users identifying harmful content they are exposed to and reducing exposure to it. Reduction of harmful content may promote well-being through two distinct mechanisms. Increased agency—the perception of control over one’s circumstances—can boost self-esteem and life satisfaction while mitigating depression and anxiety (Lee et al., 2023). Likewise, resilience, or the ability to adapt to stress and adversity, may serve as a protective factor that enhances overall well-being (Baños et al., 2017). These relationships highlight the complexity of social media’s influence on mental health, emphasizing the importance of individual differences over time. Thus, we predict:

H2: Participants in the harmful content reduction intervention will show improvement in psychological wellbeing.

H2a: Participants in the harmful content reduction intervention will show improvement in life satisfaction at week 2 and week 3 compared to control, controlling for exposure to harmful content, baseline measure of life satisfaction, self-esteem, loneliness, and time spent on Instagram (session duration and frequency of social interaction).

H2b: Participants in the harmful content reduction intervention will show lower negative self-evaluation at week 2 and week 3 compared to control, controlling for exposure to harmful content, baseline measure of negative self-evaluation, self-esteem, loneliness, and time spent on Instagram (session duration and frequency of social interaction).

H2c: Participants in the harmful content reduction intervention will show improvement in mood at week 2 and week 3 compared to the control group, controlling for baseline measure of negative mood, self-esteem, loneliness, and time spent on Instagram (session duration and frequency of social interaction).

Bi-directional Relationships Between Instagram Content and Wellbeing

A recent theoretical review has emphasized the importance of investigating bidirectional mechanisms between self-disclosure, as one of the main components of active use of social media we are interested in this study, and well-being (Luo & Hancock, 2020). In the second process of the bi-directional theoretical framework, researchers explain the effects of psychological well-being on social media as a reciprocal mechanism that influences interpersonal and intrapersonal motivations for relationship maintenance and self-expression on social media. Indeed, recent studies have identified loneliness as precursor of self-disclosure on social media (Stevic et al., 2025) as well as of passive social media use (Stevic & Matthes, 2021). As different well-being states might affect the way social media is being used to, for example, to enhance mood, we are predicting that there will be a bi-directional relationship between well-being and Instagram use. Due to reciprocal nature of relationships between social media use and well-being (Luo & Hancock, 2020) as well as loneliness (Nowland et al., 2018), we ask:

RQ1: Is there a bi-directional relationship between frequency of engagement with content on Instagram and psychological well-being?

Moreover, we are interested in how psychosocial factors and risk exposure are interconnected in a bidirectional manner based on prior research (Orben et al., 2024). Research studies highlight how individual vulnerabilities (such as mental health status) can shape one's online experiences, while at the same time, exposure to harmful content on Instagram can exacerbate or mitigate these vulnerabilities over time (Kim et al., 2024). This aligns with our approach of examining temporal

predictions to better understand the dynamic relationship between Instagram use and well-being. The importance of psychosocial factors such as social anxiety, low self-esteem, and loneliness in predicting risk exposure is crucial (Orben et al., 2024). By integrating these factors into our longitudinal models, our goal is to provide a more comprehensive understanding of risk and resilience.

We draw on the incremental theory of personality interventions (Yeager & Lee, 2021), as established in previous research on harmful online behavior (e.g., Calvete et al., 2019). This framework suggests that individuals' beliefs about their ability to change and grow play a role in how they respond to online experiences. Understanding these psychological mechanisms is crucial for identifying who is most at risk of harm and how interventions can be tailored to mitigate negative effects. We therefore aim to bidirectional effects over time between Instagram use and well-being.

RQ2: How does psychological well-being predict exposure to harmful content and harmful content reduction over time?

Interventions Design

The body of research on the impact of social media use and interactions on well-being has demonstrated both positive and negative effects, highlighting the potential to enhance user experience and mental health (Deters & Mehl, 2013; Kross, et al., 2013; Kramer, et al., 2014; Verduyn et al., 2017). Studies such as those by Kross et al. (2013) highlight the potential for social media use, particularly platforms like Facebook, to result in declines in subjective well-being. In contrast, some intervention-based research, like Deters & Mehl (2013), suggests that actively engaging with social media, such as posting status updates, can alleviate feelings of loneliness, indicating possible benefits when used intentionally.

Other scholars, like Verduyn et al. (2017), found social media can have diverse effects on users, suggesting that the relationship between social media and well-being is complex, depending on factors such as usage patterns, content exposure, and individual differences. This is supported by Sagioglou & Greitemeyer (2014), who show social media use is correlated with negative emotional consequences and explain why individuals continue to engage with these platforms despite potential mood decreases. A study of Kramer et al. (2014) on emotional contagion also demonstrates users' emotional states are significantly influenced by the emotions in the content they browse in Facebook news feeds. These studies collectively underscore the need of well-designed social media platforms to better support psychological resilience and improve mental health outcomes. In particular, they suggest design considerations around content curation and personalized context are pivotal for social media to benefit its users' well-being, providing a foundation for our proposed research on Instagram's influence on young adults' well-being.

Due to many factors occurring during Instagram use, we propose to test two different content interventions. The first will focus on user-centric quality content engagement with an emphasis on identifying and following self-selected quality content in order to understand whether promotion of positive (e.g. education, reflection, creation, entertainment) and healthy topics (e.g. nutrition, fitness, sports) enhances well-being in line with previous content interventions on Instagram (e.g., Blackstone & Russie, 2020; Depow et al., 2025). The second intervention will focus on reducing user-centric harmful content and potential risks that young adults might encounter on the platform. With instructions on how to safeguard privacy, disengage, reduce and report harmful content, we aim to help young adults have more agency over their Instagram use adapting previous intervention study on cyberbullying intervention (e.g., Kutok et al., 2021). Pre- and post-intervention surveys will assess how the intervention affects participants' psychological well-being. We will examine the persistence of intervention effects by administering follow-up surveys one week after the interventions and complement our survey-based measures with behavioral measures of Instagram use. The two interventions are separately described in the following sections.

Intervention 1: User-Centric Quality Content Engagement (Exploratory)

Participants who are randomly assigned to the *User-Centric Quality Content Intervention* condition will be encouraged to actively engage with personalized, self-identified “quality content” on their Instagram. To establish a shared understanding, we first introduce participants to a high-level definition of quality content based on the E-AIMS framework. Specifically, quality content is described as online material that promotes “minds-on” involvement, sustains focused engagement without peripheral distractions, fosters meaningful experiences connected to participants' lives, and facilitates active interaction with quality content. In line with prior research (Panahi et al., 2018; Schreurs et al., 2025), participants are told that relevant content often includes human images, artistic or scientific work, literary and cultural figures, and themes related to family, love, religion, nature, friendships, travel, and entertainment. After receiving this definition, participants are asked to reflect on their own Instagram use by responding to questions such as: “*Based on these criteria, what is the content you get value from?*” and “*What is your goal when engaging with content on Instagram?*” This reflective step introduced the positive framing of the intervention, emphasizing personally meaningful and uplifting engagement.

Participants are then instructed to browse Instagram for 10 minutes, deliberately focusing on identifying quality content in their feed. If they become distracted, they are asked to redirect their attention back toward quality posts. They are further guided to identify their most liked or most viewed posts and their most engaged-with accounts using Instagram's built-in *liked* function, uploading a screenshot of these selections to the survey (data donation). Following this, participants will reflect on what made this content valuable—such as being inspiring, relatable, or connected to close social ties—and articulate their own definition of “quality content.”

The intervention also includes a personal reflection task. Participants review their feed and identify three to five accounts or types of posts they considered as quality content for them. They are then instructed to follow five to ten new accounts that reflect their own quality content definition and values and to increase their engagement with such content by liking, saving, sharing, or commenting on user-centric quality content. Once between each wave of data collection, participants will receive reminders (via SONA) to reinforce this behavior. They will also be asked to upload a screenshot each week of one post they considered “worth their time” and to explain why it felt valuable.

Finally, after the intervention, participants will complete mood assessments after each browsing session, rating their feelings before and after exposure to quality content on a 1–10 Likert scale. At later time points (T2 and T3), when participants engage in additional self-reflection, describing a moment when they interacted with quality content in the past week, they will be asked to upload a screenshot of the post, and explaining both how it felt and what they had learned from the experience.

Intervention 2: User-Centric Harmful Content Reduction (Exploratory)

Our second intervention focuses on *User-Centric Harmful Content*. We define harmful content as content that users identify as distressful and negative for them in their Instagram feed. World Economic Forum (2023) defines harmful content as any online material that causes distress or harm to individuals and consists of different categories including contact related harms, content related harms, and conduct related harms such as threats to personal and community safety, harm to health and well-being, hate and discrimination, violation of dignity, invasion of privacy, and deception and manipulation. Rather than focusing on specific harm types, such as online hate or cyberbullying, we aim to examine a broad spectrum of harms to: (1) understand the harmful encounters and experiences on Instagram, and (2) explore the types and frequencies of harmful encounters and their associations with psychological well-being, for example, actively sought pornography might not be associated with well-being, while targeted online hate can.

In the study, we first administer a core level definition based on these categories and lead users to create their own individual level definition according to their own experiences with harmful content on Instagram. Participants assigned to the harmful content condition will be guided to focus on reducing their interaction with harmful or distressing content on Instagram. Unlike the positive framing of the quality content condition, this intervention adopts a negative focus by directing participants' attention toward content they regretted seeing, engaging with or found upsetting. To begin, participants are instructed to browse Instagram for ten minutes while paying close attention to posts they perceive as harmful based on the provided definition.

As part of this process, participants are asked to upload screenshots of posts that they considered harmful, upsetting, or toxic. They then categorized these posts and generated a personal definition

of “harmful content.” Examples might include unrealistic body images, negative or sensationalized news, and influencer content that lowers their self-esteem, though participants will be encouraged to expand these categories to fit their own experiences.

Following this identification task, participants receive guidance on reappraisal strategies designed to reduce their exposure to harmful content. These included practical steps such as unfollowing or muting problematic accounts, using the “Report” or “Not Interested” functions, and enabling Instagram’s “Favorites Only” mode when available.

At the conclusion of intervention, participants will be asked to rate their feelings before and after exposure to self-identified harmful content using a 1–10 Likert mood scale. They will also be asked whether they had avoided or removed harmful content during the intervention. At subsequent surveys, participants will be asked to engage in self-reflection. For example, at week 2, they will be asked to describe a moment during the past week when they avoided harmful content, to reflect on how this experience felt, and to explain what they learned from the process.

Finally, we ask:

RQ3: To what extent do the two intervention groups differ in their impact on psychological well-being?

Methods

We will conduct a three-wave longitudinal study with one-week intervals between each wave among young adults. One of the main reasons for keeping short-term intervals is the success rate of the interventions. Therefore, we have decided to utilize a shorter time interval between the three surveys and rely on a two-week time interval based on earlier intervention studies on Instagram harmful experiences (Kim et al., 2024). Dormann & Griffin (2015) demonstrated that shorter time lags between surveys are more optimal to show changes in psychological outcomes than longer time intervals between the waves. Additionally, due to changes in app use, we aim to keep most of our participants eligible for continuation of the survey according to their status as Instagram users. This research has been approved by the Institutional Review Board at Stanford University (Protocol # IRB-78354). All participants will provide informed consent online through the Qualtrics survey platform. We aim to recruit young adult Instagram users (ages between 18 and 25 years) via a convenient snowball sampling method through university courses and online recruitment platforms. We expect a 40%–50% drop-out rate based on previous panel studies in young adult populations (Arendt et al. 2019; Geusens & Beullens, 2021; Koban et al., 2023). We plan to over recruit participants in the first survey aiming for $N = 500$ in order to reach 300 young adults in the third survey wave. All participants recruited from university participants’ pools will receive credit for completing all three surveys and tasks. Based on our a priori power analysis and our current experimental setup, we need at least 80 participants per group (a total of 240) to detect an

effect size of $d = 0.3$ to reach a minimum of .80 power. The script including simulated data and power analysis is included as Appendix 1.

In the survey questionnaire, we will include two attention check questions to assess participant attentiveness and exclude responses that fail these checks. Additionally, we will clean the data by identifying and removing participants who complete the survey too quickly (indicating insufficient engagement) or who do not pass manipulation checks. Furthermore, to enhance the robustness of our results, we will conduct sensitivity analyses to assess the stability of our findings across different analytical approaches and potential variations in the dataset, e.g., we will rely on different model specifications using different estimation techniques. These steps will help ensure that our conclusions are based on high-quality data and are not unduly influenced by methodological artifacts.

We will conduct panel surveys asking users about their psychological well-being at three time points, with a one-week interval in between each wave. In parallel, we plan to collect users' actual behavioral Instagram data and combine it with their self-reported experiences for analysis (see Figure 1). We will collect well-being measurements including physical and psychological indicators at all three time points to allow for testing over time changes. All self-reported measures in the surveys will be based on the validated scales used in communication science research. Variables will include established brief scales from the literature including life events scale (Dohrenwend et al., 1978), Positive and Negative Affect (PANAS; Watson, Clark, & Tellegen, 1988), life satisfaction scale (Diener et al., 1985), loneliness (UCLA scale by Hughes, Waite, Hawkey, & Cacioppo, 2004), and self-esteem (Rosenberg et al., 1979). Each study wave will include a self-report questionnaire and intervention. Key conditions include an intervention that includes a task of identifying and following quality content vs. identifying and reducing harmful content and a control group that receives no intervention. At the end of each survey, we design an intervention as a task to be completed and plan to test its effectiveness in the follow-up survey.

In the first survey (T1), we will randomly assign participants either to the user-centric quality content intervention, user-centric harmful content intervention or control group with no intervention. In the two treatment groups, we are asking participants to identify and engage more with quality content or identify and engage less with harmful content. The success of the intervention will be monitored through manipulation check questions asking about identifying the quality content accounts and providing a screenshot of the post showing that they have seen it and/or engaged with the account or post.

In the second survey (T2), participants in the two content intervention treatments will continue with the same intervention and identify what has changed in the past week in terms of content exposure and associated feelings towards quality or harmful content. We will ask participants to engage with quality content again more actively through exposure, sharing, commenting, liking, or following accounts. We will measure if they see and engage with quality content by asking participants to upload screenshots. Similarly, we will ask participants to engage less with harmful

content and unfollow, report, mute or remove such content. We will not directly measure Instagram content, as this information is not accessible to researchers. Instead, we will assess perceived exposure to harmful content by asking participants to upload screenshots and report on their experiences.

Measures

Life satisfaction. We will assess life satisfaction pre- and post-intervention with one item from Harvard flourishing scale, asking participants on a 7-point scale (1 = Not satisfied at all, 7 = Completely satisfied), how satisfied are you with life as a whole these days?

Self-Esteem. We will assess state self-esteem pre-intervention and post-intervention, asking participants to indicate on a 7-point scale (1 = Really bad, 7 = Really good) how they feel about themselves in the present moment (Burnell et al., 2020).

Mood. We will assess mood with the brief positive and negative affect scale (PANAS; Watson et al., 1988) including five positive items such as experiencing feelings of enthusiasm, inspiration, excitement, and five negative emotions such as being afraid, angry and distressed.

Self-Perceptions. We will assess self-perceptions pre- and post-intervention with two items of interpersonal negativity including feelings of envy and jealousy, and positive self-perceptions including feelings of confidence and being valued based on Burnell et al. (2020). Participants will rate how they feel on a 7-point scale (1 = I do not feel this way right now, 7 = I definitely feel this way right now).

Loneliness. We will assess loneliness with three items from a brief UCLA scale, asking participants if they lack companionship, feel left out and isolated on a 5-point scale (1 = strongly disagree to 5 = strongly agree).

Social Comparison. We assess social comparison with one item from the Depow et al. (2025) study, asking participants on a 5-point scale ranging from very rarely or never (1) to very often or always (5) to indicate how often did you compare yourself to others? We will also utilize social comparison as a moderator in exploratory analysis since its role is established in previous research (Vogel et al., 2015; de Vries et al., 2018).

Design and Analytical Approach

To address potential attrition from the surveys, we will implement several strategies to mitigate dropouts and ensure data quality. First, we will enhance participant engagement by offering self-selection of influencers based on different topics. Additionally, we will send reminder notifications during the two-week time interval and offer incentives for survey completion to reduce participant loss over time. To address potential attrition from the intervention condition, we will implement several strategies to mitigate dropouts and ensure data quality. First, we will enhance participant

engagement by offering self-selection of influencers based on different topics. Additionally, we will send reminder notifications during the two-week time interval and offer incentives for survey completion to reduce participant loss over time. Due to some expected attrition rates we will evaluate whether attrition is systematic. We will conduct chi-square analysis and independent samples t-tests on key demographic and outcome variables (e.g., age, gender, well-being indicators). These analyses will help identify whether participants who dropped out differ significantly from those who remained in the study. If we detect systematic differences, we will further consider implementing statistical adjustments—such as weighting or multiple imputation—to preserve the validity of our results.

The study employs a three-group experimental design to examine the effects of distinct Instagram use interventions on psychological well-being. Participants are randomly assigned to one of three conditions: (1) the *User-Centered Quality Content Intervention*, which encourages engagement with personally meaningful and uplifting content; (2) the *Harmful Content Reduction Intervention*, which directs attention toward identifying and reducing exposure to harmful or distressing content; or (3) a *control group*, in which participants use Instagram as usual without structured guidance regarding content. Each condition is designed to simulate real-world browsing behaviors while systematically manipulating the type of content participants attend to, reflect upon, and share through screenshots. The interventions are administered longitudinally across three time points, enabling the examination of both immediate and sustained effects on psychological well-being outcomes.

To evaluate the intervention effects, we follow a longitudinal ANCOVA framework, controlling for relevant baseline variables and lagged measures. Specifically, life satisfaction, self-esteem, loneliness, time spent on Instagram (session duration), and levels of engagement metrics (e.g., commenting, liking, sharing, posting, changing features) are entered as covariates. This analytic approach allows us to isolate the unique contribution of the interventions while accounting for individual differences in baseline well-being and platform use.

We implement the analysis as a linear mixed-effects model with random intercepts to account for repeated measures nested within participants. This model specification adjusts for within-person correlation across time and captures individual differences in baseline outcome levels. Fixed effects for intervention condition, time, and their interaction were included, along with baseline covariates (life satisfaction, self-esteem, loneliness, Instagram use). This specification allows robust estimation of intervention-related differences at post-baseline time points while accommodating heterogeneity in participants' responses.

Considering the effect sizes of the predictors, our results must account for the stability of pathways among variables. The strong autoregressive effects we anticipate suggest that changes in well-being variables over time are minimal. As a result, controlling for stability significantly reduces variance. Consequently, given the high stability of the outcomes, even small effects from other predictors remain meaningful (Adachi & Willoughby, 2015). Lastly, to exploratively study potential

user differences, we will rely on latent class analysis (Foerster & Rösli, 2017) as well as moderation analysis including social comparison and metrics from behavioral Instagram data regarding use of control features and privacy set up.

The analysis aims to explore how different types of engagement and user-generated activities on Instagram are associated with psychological well-being outcomes. Dependent variables (measured at T1, baseline survey) will include psychological well-being denoting positive outcomes in terms of self-esteem and life satisfaction and negative outcomes in terms of loneliness and negative self-evaluation. Predictors (measured from T0 to T1) will include (1) time spent (session duration) that is the total duration a user spends on Instagram within the month leading up to the first survey; (2) close-tie content view defined as the frequency of views or duration of time users spend viewing content posted by close connections, close-tie content is defined as content from users who follow each other or content designated by the poster to be seen only by close friends; (3) public content view defined as the frequency of views or duration spent viewing content that is publicly accessible including ads; (4) user-generated content including post creation (the number of posts a user creates within the period from T0 to T1), comment creation (the number of comments a user makes during the specified one-month period), like activity (the total number of times a user clicks the 'like' button during the same timeframe); (5) harmful content exposure and harmful encounters. Control variables will include age, gender, session duration, privacy settings, and life events. The overview of study design and detailed approach based on hypotheses and research questions is presented in Table 1.

Intervention 1: Quality Content Engagement

Testing the first intervention will include dependent variables that are well-being scores collected at T2 and T3 from indicators of positive and negative affect, self-esteem, depressive symptoms, anxiety, and life satisfaction as well as behavioral engagement on Instagram in terms of time spent (session duration), interactive behavior (commenting, liking, interacting), passive browsing behavior (close-tie content views, public content views). Independent variable is the intervention treatment at T1 coded as a binary variable, 1 if the user is assigned to the treatment group 1, otherwise 0. Control Variables: session duration, baseline measures of wellbeing.

Intervention 2: Harmful Content Reduction

Dependent variables will include well-being scores collected at T2 and T3: positive and negative affect, self-esteem, self-evaluation, and life satisfaction as well as interactive and browsing engagement on Instagram including variables listed above and use of safety and control features (privacy, blocking, reporting) on Instagram. An independent variable is harm reduction intervention.

Instagram Content Measures and Analysis

We will not directly measure Instagram content, as this information is not accessible to researchers. Instead, we will assess content by asking participants to upload screenshots of (1) self-identified quality content posts; (2) self-identified harmful content posts; (3) most liked posts; (4) first three posts of their Instagram news feed; (5) profile accounts of newly followed accounts. These images will be analyzed to identify the main topics and content elements appearing in users' feeds and compared to the user-centric definition of quality content and harmful content. Our approach will include manual content analysis. First, we will conduct a qualitative review of the screenshots to identify key themes and content categories to create a codebook and de-identify any user information. This process ensures that we capture nuanced aspects of the content that may not be easily recognized with other methods of analysis. Second, we will code the screenshots. After intercoder reliability tests, coding process will include systematic categorization of the visual elements and classification of the content that participants are exposed to in their news feeds in terms of quality and harmful content.

Instagram Data Access Pilot Transparency Statement

For a fuller view of this project's process through the Instagram Data Access Pilot, supporting documents are openly available here: <https://osf.io/thuqp/>

Supporting Information

Supporting information files related to this work can be found here: <https://osf.io/5r2yg>

References

- Agha, Z., Badillo-Urquiola, K., & Wisniewski, P. J. (2023). "Strike at the Root": Co-designing real-time social media interventions for adolescent online risk prevention. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1–32.
- Alonzo, R., Hussain, J., Stranges, S., & Anderson, K. K. (2021). Interplay between social media use, sleep quality, and mental health in youth: A systematic review. *Sleep Medicine Reviews*, 56, 101414.
- Alsoubai, A., Razi, A., Agha, Z., Ali, S., Stringhini, G., De Choudhury, M., & Wisniewski, P. J. (2024). Profiling the offline and online risk experiences of youth to develop targeted interventions for online safety. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), 1–37.

Arendt, F., Scherr, S., & Romer, D. (2019). Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults. *New Media & Society*, 21(11–12), 2422–2442.

<https://doi.org/10.1177/1461444819850106>

Assembly Committee on Education (2024). AB 3216 SUBJECT: Pupils: use of smartphones. Retrieved from <https://aedn.assembly.ca.gov/system/files/2024-04/ab-3216.pdf>

Baños, R. M., Etchemendy, E., Mira, A., Riva, G., Gaggioli, A., & Botella, C. (2017). Online positive interventions to promote well-being and resilience in the adolescent population: A narrative review. *Frontiers in Psychiatry*, 8, 10.

Beyens I., Pouwels J. L., van Driel I. I., Keijsers L., Valkenburg P.M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports*, 10(1), 1–11.

<https://doi.org/10.1038/s41598-020-67727-7>

Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L., & Valkenburg, P.M. (2024). Social Media Use and Adolescents' Well-Being: Developing a Typology of Person-Specific Effect Patterns. *Communication Research*, 51(6), 691–716. <https://doi.org/10.1177/00936502211038196>

Blackstone, S., & Russie, C. (2020). Using Instagram to deliver a theory-based wellness intervention designed to promote behavior change. *European Journal of Public Health*, 30(Supplement_5), ckaa166–008. <https://doi.org/10.1093/eurpub/ckaa166.008>

Bonnie, R. J., Stroud, C., Breiner, H., Committee on Improving the Health, S., & National Research Council. (2015). Young adults in the 21st century. In *Investing in the health and well-being of young adults*. National Academies Press (US).

Burke, M. & Kraut, R. E. (2016). The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength, *Journal of Computer-Mediated Communication*, 21(4), 265–281, <https://doi.org/10.1111/jcc4.12162>

Burnell, K., George, M. J., & Underwood, M. K. (2020). Browsing different Instagram profiles and associations with psychological well-being. *Frontiers in Human Dynamics*, 2, 585518. doi.org/10.3389/fhumd.2020.585518

Capio, C. M., Sit, C. H. P., Abernethy, B. (2014). Physical Well-Being. In: Michalos, A. C. (eds) *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_2166

Cash, T. F., & Fleming, E. C. (2002). The impact of body image experiences: development of the body image quality of life inventory. *International Journal of Eating Disorders*, 31(4), 455–460. <https://doi.org/10.1002/eat.10033>

- Chan M. (2015). Multimodal connectedness and quality of life: Examining the influences of technology adoption and interpersonal communication on well-being across the life span. *Journal of Computer-Mediated Communication*, 20(1), 3–18. <https://doi.org/10.1111/jcc4.12089>
- Chan M. (2018). Mobile-mediated multimodal communications, relationship quality and subjective well-being: An analysis of smartphone use from a life course perspective. *Computers in Human Behavior*, 87, 254–262. <https://doi.org/10.1016/j.chb.2018.05.027>
- Depow, G. J., Oldemburgo de Mello, V., & Inzlicht, M. (2025). A positive empathy intervention to improve well-being on Instagram. *Emotion*. Advance online publication. <https://doi.org/10.1037/emo0001489>
- Deters, F. G., & Mehl, M. R. (2013). Does posting Facebook status updates increase or decrease loneliness? An online social networking experiment. *Social psychological and personality science*, 4(5), 579–586.
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
- Ding, D., Tan, X., Sun, L., Zeng, X., & Yan, Z. (2024). Does empathy lead to depression? A three-wave random intercept cross-lagged panel analysis in emerging adults. *Current Psychology*, 43(12), 10971–10978. <https://doi.org/10.1007/s12144-023-05224-6>
- Dohrenwend, B. S., Askenasy, A. R., Krasnoff, L., & Dohrenwend, B. P. (1978). Exemplification of a method for scaling life events: The PERI Life Events Scale. *Journal of Health and Social Behavior*, 205–229. <https://doi.org/10.2307/2136536>
- Douglas, K. D., Smith, K. K., Stewar, M. W., Walker, J., Mena, L., Zhang, L. (2023). Exploring Parents' Intentions to Monitor and Mediate Adolescent Social Media Use and Implications for School Nurses. *The Journal of School Nursing*, 39(3), 248–261. doi:10.1177/1059840520983286
- Evans, G. W., & English, K. (2002). The Environment of Poverty: Multiple Stressor Exposure, Psychophysiological Stress, and Socioemotional Adjustment. *Child Development*, 73(4), 1238–1248. <https://doi.org/10.1111/1467-8624.00469>
- Falkenström, F., Solomonov, N., & Rubel, J. A. (2022). How to model and interpret cross-lagged effects in psychotherapy mechanisms of change research: A comparison of multilevel and structural equation models. *Journal of Consulting and Clinical Psychology*, 90(5), 446. <https://dx.doi.org/10.1037/ccp0000727>
- Foerster, M., & Rösli, M. (2017). A latent class analysis on adolescents media use and associations with health related quality of life. *Computers in Human Behavior*, 71, 266–274. <https://doi.org/10.1016/j.chb.2017.02.015>

Geusens, F., & Beullens, K. (2021). Triple spirals? A three-wave panel study on the longitudinal associations between social media use and young individuals' alcohol consumption. *Media Psychology*, 24(6), 766–791. <https://doi.org/10.1080/15213269.2020.1804404>

Gualdi-Russo, E., Rinaldo, N., & Zaccagni, L. (2022). Physical activity and body image perception in adolescents: a systematic review. *International Journal of Environmental Research and Public Health*, 19(20), 13190.

Haidt J. (2024). The anxious generation: how the great rewiring of childhood is causing an epidemic of mental illness. London: Penguin Press.

Hall, J. A. (2020). Relating through technology: Advances in personal relationships.

Hancock, J., Liu, S. X., Luo, M., & Mieczkowski, H. (2022). Psychological Well-Being and Social Media Use: A Meta-Analysis of Associations between Social Media Use and Depression, Anxiety, Loneliness, Eudaimonic, Hedonic and Social Well-Being. <http://dx.doi.org/10.2139/ssrn.4053961>

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102–116. <https://doi.org/10.1037/a0038889>

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in “Educational” Apps: Lessons From the Science of Learning. *Psychological Science in the Public Interest*, 16(1), 3–34. <https://doi.org/10.1177/1529100615569721>

Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on Aging*, 26(6), 655–672. <https://doi.org/10.1177/0164027504268574>

Jung, J., Barron, D., Lee, Y. A., & Swami, V. (2022). Social media usage and body image: Examining the mediating roles of internalization of appearance ideals and social comparisons in young women. *Computers in Human Behavior*, 135, 107357. <https://doi.org/10.1016/j.chb.2022.107357>

Keipi, T., Näsi, M., Oksanen, A., & Räsänen, P. (2017). *Online Hate and Harmful Content: Cross-national perspectives*. Routledge.

Keles B., McCrae N., Grealish A. (2020). A systematic review: The influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*, 25(1), 79–93. <https://doi.org/10.1080/02673843.2019.1590851>

Kim, S., Razi, A., Alsoubai, A., Wisniewski, P. J., & De Choudhury, M. (2024, May). Assessing the Impact of Online Harassment on Youth Mental Health in Private Networked Spaces. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 18, pp. 826–838).

Kramer, A.D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.

Kreski, N. T., Chen, Q., Olfson, M., Cerdá, M., Martins, S. S., Mauro, P.M., ... & Keyes, K. M. (2022). Experiences of online bullying and offline violence-related behaviors among a nationally representative sample of US adolescents, 2011 to 2019. *Journal of School Health*, 92(4), 376–386.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., ... & Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. *PloS One*, 8(8), e69841.

Kutok, E. R., Dunsiger, S., Patena, J. V., Nugent, N. R., Riese, A., Rosen, R. K., & Ranney, M. L. (2021). A cyberbullying media-based prevention intervention for adolescents on instagram: pilot randomized controlled trial. *JMIR Mental Health*, 8(9), e26029.

Landesman, R., Yoon, J., Kim, J., Munoz Lopez, D. E., Magis-Weinberg, L., Hiniker, A., & Davis, K. (2024, June). "I just don't care enough to be interested": Teens' moment-by-moment experiences on Instagram. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference* (pp. 14–29).

Lee, C., Gligorić, K., Kalluri, P. R., Harrington, M., Durmus, E., Sanchez, K. L., ... & Eberhardt, J. L. (2024). People who share encounters with racism are silenced online by humans and machines, but a guideline-reframing intervention holds promise. *Proceedings of the National Academy of Sciences*, 121(38), e2322764121.

Lee, A. Y., & Hancock, J. T. (2024). Social media mindsets: A new approach to understanding social media use and psychological well-being. *Journal of Computer-Mediated Communication*, 29(1), zmad048.

Lee, J. A., Sudarshan, S., Sussman, K. L., Bright, L. F., & Eastin, M. S. (2022). Why are consumers following social media influencers on Instagram? Exploration of consumers' motives for following influencers and the role of materialism. *International Journal of Advertising*, 41(1), 78–100.
<https://doi.org/10.1080/02650487.2021.1964226>

Mackson, S. B., Brochu, P.M., & Schneider, B. A. (2019). Instagram: Friend or foe? The application's association with psychological well-being. *New Media & Society*, 21(10), 2160–2182.
<https://doi.org/10.1177/1461444819840021>

Marciano, L., Lin, J., Sato, T., Saboor, S., & Viswanath, K. (2024). Does social media use make us happy? A meta-analysis on social media and positive well-being outcomes. *SSM - Mental Health*, 6, 100331. <https://doi.org/10.1016/j.ssmmh.2024.100331>

Meier, A., Gilbert, A., Börner, S., & Possler, D. (2020). Instagram inspiration: How upward comparison on social network sites can contribute to well-being. *Journal of Communication*, 70(5), 721–743.

National Academies of Sciences, Engineering, and Medicine (2024). Social media and adolescent health. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27396>.

Nowland, R., Necka, E. A., & Cacioppo, J. T. (2018). Loneliness and Social Internet Use: Pathways to Reconnection in a Digital World? *Perspectives on Psychological Science*, 13(1), 70–87. <https://doi.org/10.1177/1745691617713052>

Maes, C., & Vandenbosch, L. (2022). Adolescent girls' Instagram and TikTok use: Examining relations with body image-related constructs over time using random intercept cross-lagged panel models. *Body image*, 41, 453–459. <https://doi.org/10.1016/j.bodyim.2022.04.015>

McNair, D. M. (1971). Manual for the Profile of Mood States. *Educational and Industrial Testing Services*.

Murthy, V. (2024). Surgeon General Issues New Advisory About Effects Social Media Use Has on Youth Mental Health. Retrieved from <https://www.hhs.gov/about/news/2023/05/23/surgeon-general-issues-new-advisory-about-effects-social-media-use-has-youth-mental-health.html>

Koban, K., Stevic, A., & Matthes, J. (2023). A tale of two concepts: differential temporal predictions of habitual and compulsive social media use concerning connection overload and sleep quality. *Journal of Computer-Mediated Communication*, 28(2), zmac040. <https://doi.org/10.1093/jcmc/zmac040>

Lee, A. Y., & Hancock, J. T. (2024). Social media mindsets: A new approach to understanding social media use and psychological well-being. *Journal of Computer-Mediated Communication*, 29(1), zmad048. <https://doi.org/10.1093/jcmc/zmad048>

Lee, H., Lee, H., Kim, Y. *et al.* Reciprocal relationship between multicultural adolescents' depression and life satisfaction: a random intercept cross-lagged panel model for 3-wave panel data. *Applied Research Quality Life*, 17, 2353–2367 (2022). <https://doi.org/10.1007/s11482-021-10032-w>

Luo, M., & Hancock, J. T. (2020). Self-disclosure and social media: Motivations, mechanisms and psychological well-being. *Current Opinion in Psychology*, 31, 110–115. <https://doi.org/10.1016/j.copsyc.2019.08.019>

Orben, A., & Przybylski, A. K. (2019). Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. *Psychological Science*, 30(5), 682–696. <https://doi.org/10.1177/0956797619830329>

Raychoudhury, P. (2021). What Our Research Really Says About Teen Well-Being and Instagram. Retrieved from <https://about.fb.com/news/2021/09/research-teen-well-being-and-instagram/>

Rosenberg, M. (1979). *Conceiving the Self*. New York: Basic Books.

Sagioglou, C., & Greitemeyer, T. (2014). Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it. *Computers in Human Behavior*, 35, 359–363.

Saiphoo, A. N., Halevi, L. D., & Vahedi, Z. (2020). Social networking site use and self-esteem: A meta-analytic review. *Personality and Individual Differences*, 153, 109639.

Sherlock, M., & Wagstaff, D. L. (2019). Exploring the relationship between frequency of Instagram use, exposure to idealized images, and psychological well-being in women. *Psychology of Popular Media Culture*, 8(4), 482–490. <https://doi.org/10.1037/ppm0000182>

Schmuck D., Karsay K., Matthes J., Stevic A. (2019). "Looking up and feeling down." The influence of mobile social networking site use on upward social comparison, self-esteem, and well-being of adult smartphone users. *Telematics and Informatics*, 42, Article 101240. <https://doi.org/10.1016/j.tele.2019.101240>

Tynes, B. M., Lozada, F. T., Smith, N. A., & Stewart, A.M. (2020). Trajectories of Online Racial Discrimination and Psychological Functioning Among African American and Latinx Adolescents. *Child Development*, 91(3), 902–918. <https://doi.org/10.1111/cdev.13350>

Valkenburg, P.M., van Driel, I. I., & Beyens, I. (2022). The associations of active and passive social media use with well-being: A critical scoping review. *New Media & Society*, 24(2), 530–549.

Verduyn, P., Gugushvili, N., & Kross, E. (2022). Do Social Networking Sites Influence Well-Being? The Extended Active-Passive Model. *Current Directions in Psychological Science*, 31(1), 62–68. <https://doi.org/10.1177/09637214211053637>

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, 11(1), 274–302.

Vuorre, M., & Przybylski, A. K. (2024). Global Well-Being and Mental Health in the Internet Age. *Clinical Psychological Science*, 12(5), 917–935. <https://doi.org/10.1177/21677026231207791>

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>

Figure 1.

Research Design

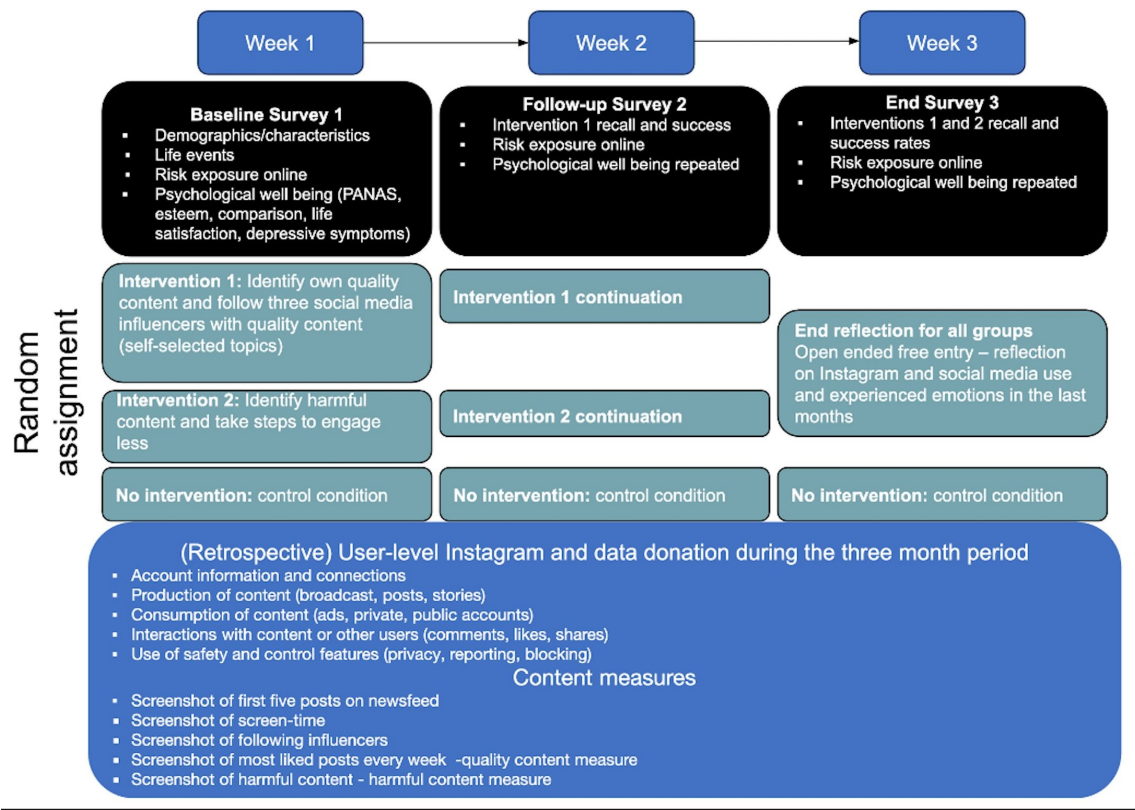


Table 1.

Study Design Overview

Question	Hypothesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
How does exposure and engagement with quality content relate to wellbeing over time?	<i>H1:</i> Participants in the quality content intervention will show improvement in psychological wellbeing.					

	<i>H1a:</i> Participants in the quality content intervention will show improvement in life satisfaction at week 2 and week 3 compared to the control group, controlling for baseline measure of life satisfaction, and time spent on Instagram (session duration and frequency of social interaction). <i>H1b:</i> Participants in the quality content intervention will show improvement in mood at week 2 and week 3 compared to the control group, controlling for baseline measure of positive mood, and time spent on Instagram (session duration and					
--	--	--	--	--	--	--

	frequency of social interaction). <i>H1c:</i> Participants in the quality content intervention will show lower negative self-evaluation and higher positive self-evaluation at week 2 and week 3 compared to the control group, controlling for baseline measure of negative self-evaluation, depression, self-esteem, loneliness, and time spent on Instagram (session duration and frequency of social interaction).	We plan to over recruit participants in the first survey aiming for N = 500 in order to reach 300 young adults in the third survey wave. All participants recruited from university participants' pools will receive credit for completing all three surveys and tasks. Based on our a priori power analysis and our current experimental setup, we need at least 80 participants per group (a total of 240) to detect an effect size of $d = 0.3$ to reach a minimum of .80 power. The script including simulated data and power analysis is included as Appendix 1.	Longitudinal ANCOVA linear mixed-effects model with random intercepts to account for repeated measures nested within participants; lagged well-being measure (the multilevel regression analysis controls for an individual's previous level of well-being)	Power analysis based on simulated data	If we find support for H1, H1a, H1b, H1c at $p < .05$, we will conclude that the causes of overall increases in psychological wellbeing including mood, life satisfaction, and self-evaluation are due to quality content exposure and engagement . If we find negative evidence for H1, we will conclude the opposite: that the quality content exposure and engagement decreases psychological well-being including mood, life satisfaction, and self-evaluation If we find no significant relationship, we will conclude that quality	
--	---	---	--	---	---	--

					content over time does not relate to psychological well-being including mood, life satisfaction, and self-evaluation.	
How does harmful content reduction relate to wellbeing over time?	<i>H2:</i> Participants in the harmful content reduction intervention will show improvement in psychological wellbeing. <i>H2a:</i> Participants in the harmful content reduction intervention will show improvement in life satisfaction at week 2 and week 3 compared to control, controlling for exposure to harmful content, baseline measure of life satisfaction, self-esteem, loneliness, and time spent on					

	Instagram (session duration and frequency of social interaction). <i>H2b:</i> Participants in the harmful content reduction intervention will show lower negative self-evaluation at week 2 and week 3 compared to control, controlling for exposure to harmful content, baseline measure of negative self-evaluation, self-esteem, loneliness, and time spent on Instagram (session duration and frequency of social interaction). <i>H2c:</i> Participants in the harmful content reduction intervention will show					
--	---	--	--	--	--	--

	t in mood at week 2 and week 3 compared to the control group, controlling for baseline measure of negative mood, self-esteem, loneliness, and time spent on Instagram (session duration and frequency of social interaction).	We plan to over recruit participants in the first survey aiming for N = 500 in order to reach 300 young adults in the third survey wave. All participants recruited from university participants' pools will receive credit for completing all three surveys and tasks. Based on our a priori power analysis and our current experimental setup, we need at least 80 participants per group (a total of 240) to detect an effect size of $d = 0.3$ to reach a minimum of .80 power. The script including simulated data and power analysis is included as Appendix 1.	Longitudinal ANCOVA linear mixed-effects model with random intercepts to account for repeated measures nested within participants; lagged well-being measure (the multilevel regression analysis controls for an individual's previous level of well-being).	Power analysis based on simulated data	If we find support for H2, H2a, H2b, H2c at $p < .05$, we will conclude that the causes of overall increases in psychological wellbeing including mood, life satisfaction, and self-evaluation are due to harmful content reduction. If we find negative evidence for H1, we will conclude the opposite: that the harmful content reduction decreases psychological well-being including mood, life satisfaction, and self-evaluation. If we find no significant relationship, we will conclude that harmful content reduction does not relate to	
--	--	---	---	---	--	--

					I well-being including mood, life satisfaction, and self-evaluation.	
RQ1: Is there a bi-directional relationship between frequency of engagement with content on Instagram and psychological well-being?		We plan to over recruit participants in the first survey aiming for N = 500 in order to reach 300 young adults in the third survey wave. All participants recruited from university participants' pools will receive credit for completing all three surveys and tasks. Based on our a priori power analysis and our current experimental setup, we need at least 80 participants per group (a total of 240) to detect an effect size of $d = 0.3$ to reach a minimum of .80 power. The script				

		including simulated data and power analysis is included as Appendix 1.	Longitudinal ANCOVA linear mixed-effects model with random intercepts to account for repeated measures nested within participants; lagged well-being measure (the multilevel regression analysis controls for an individual's previous level of well-being).	Power analysis based on simulated data	For RQ1, if we find a positive relationship that higher well-being predicts more active engagement over time at $p < .05$, we will conclude it is because users with higher psychological well-being may engage more actively, using Instagram for meaningful interactions intentionally benefitting them. If higher well-being relates to less engagement we will conclude that those users with high well-being might prioritize offline interactions, use Instagram more intentionally, or limit screen time. If we find that lower	
--	--	--	--	--	---	--

					well-being is related to more engagement we will conclude that users with poor well-being may seek social support, distraction, or emotional relief, possibly leading to excessive use. If we find that lower well-being relates to less engagement we will conclude that psychological distress may reduce motivation to interact online, leading to withdrawal or passive browsing. If we find no significant relationship, we will conclude that active engagement on Instagram does not relate to	
--	--	--	--	--	--	--

					I wellbeing, i.e. psychologica I wellbeing is not predicting engagement .	
RQ2: How does psychologica I well-being predict exposure to harmful content and harmful content reduction over time?		We plan to over recruit participants in the first survey aiming for N = 500 in order to reach 300 young adults in the third survey wave. All participants recruited from university participants' pools will receive credit for completing all three surveys and tasks. Based on our a priori power analysis and our current experimenta I setup, we need at least 80 participants per group (a total of 240) to detect an effect size of d = 0.3 to reach a minimum				

		of.80 power. The script including simulated data and power analysis is included as Appendix 1.	Longitudinal ANCOVA linear mixed-effects model with random intercepts to account for repeated measures nested within participants; lagged well-being measure (the multilevel regression analysis controls for an individual's previous level of well-being).	Power analysis based on simulated data	For RQ2, if we find that higher well-being predicts more exposure to harmful content and harmful content reduction over time at $p<.05$, we will conclude it is because users with higher psychological well-being may engage more actively. If higher well-being relates to less exposure to harmful content and harmful content reduction we will conclude that those users with high well-being might prioritize offline interactions, use Instagram more intentionally, or limit screen time. If we find that lower well-being is	
--	--	--	--	--	--	--

					related to more exposure to harmful content and harmful content reduction we will conclude that users with poor well-being may search for more content, distraction, or emotional relief, possibly leading to excessive use. If we find that lower well-being relates to less exposure to harmful content and harmful content reduction we will conclude that psychological distress may reduce motivation to be on Instagram. If we find no significant relationship, we will conclude that exposure to harmful content and	
--	--	--	--	--	---	--

					harmful content reduction on Instagram does not relate to psychological wellbeing, i.e. psychological wellbeing is not predicting reduction.	
RQ3: To what extent do the two intervention groups differ in their impact on psychological well-being?		We plan to over recruit participants in the first survey aiming for N = 500 in order to reach 300 young adults in the third survey wave. All participants recruited from university participants' pools will receive credit for completing all three surveys and tasks. Based on our a priori power analysis and our current experimental setup, we need at least 80 participants per group (a				

		total of 240) to detect an effect size of $d = 0.3$ to reach a minimum of .80 power. The script including simulated data and power analysis is included as Appendix 1.	Longitudinal ANCOVA linear mixed-effects model with random intercepts to account for repeated measures nested within participants; lagged well-being measure (the multilevel regression analysis controls for an individual's previous level of well-being).	Power analysis based on simulated data	For RQ3, if higher significant effect size is shown for the quality content engagement intervention than for harmful content reduction intervention, we conclude that exposure to and engagement with quality content positively impacts users' psychological well-being, possibly by providing inspiration, positivity, or meaningful engagement . If higher significant effect size is shown for the harmful content reduction intervention than for quality content engagement intervention, we conclude that exposure to harmful	
--	--	--	--	--	--	--

					content reduction might increase positive mood, esteem and increase well-being. If there are no differences between groups, we conclude that the interventions may not be strong enough to influence well-being.	
--	--	--	--	--	---	--

Declarations

Ethics

The research has received an IRB approval (protocol number 78354) from Stanford University.

Competing Interests

The authors declare that no conflicts of interest exist.

Funding

This research did not receive funding.

Author Contributions

AS: Conceptualization, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing YCW: Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing DDSV: Conceptualization, Writing – original draft, Writing – review & editing SXL: Conceptualization, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing JH: Conceptualization, Methodology, Project administration, Resources, Supervision