Computers in Human Behavior 160 (2024) 108360



Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh



Real-time incentivizing survey completion with game-based rewards in experience sampling research may increase data quantity, but reduces data quality☆

Egon Dejonckheere [a,b,*], Stijn Verdonck [a], Joren Andries [a], Natalie Röhrig [a], Maarten Piot [a], Ghijs Kilani [a], Merijn Mestdagh [a]

[a] KU Leuven, Faculty of Psychology and Educational Sciences, Belgium
[b] Tilburg University, Department of Medical and Clinical Psychology, The Netherlands

ARTICLE INFO

Handling editor: Nicolae Nistor

Keywords:
Ecological momentary assessment
Daily life research
Study compliance
Careless responding
Response streak

ABSTRACT

Experience sampling methodology (ESM) requires participants to repeatedly rate various psychological states in everyday life. This can be burdensome, leading to frequently skipped or missed assessments that undermine the ecological validity of collected data. Financially compensating participants may counter suboptimal response rates, but monetary rewards are not always feasible or ethical, and are unfitting to counter non-response in the moment. Here, we investigated the potential of a gamified momentary reward strategy to encourage survey completion. In a between-person experiment, we randomly assigned 193 participants to take part in a standard (i.e., no momentary incentives) or gamified ESM protocol. Upon completing a survey in the gamified condition, participants immediately received some virtual coins to purchase various in-app rewards (fun facts or personalized graphs). Gamifying momentary survey completion resulted in higher response rates and shorter latencies, but only for participants who frequently bought rewards in the coin store. Paradoxically, momentary gamification also caused participants' responses to be slightly more unreliable. Completion times did not differ. The reward schemes of future ESM studies should focus on both response quantity and quality. To further explore the potential of momentary incentive strategies, we provide open access to customizable gamification building blocks within our ESM platform, m-Path.

To investigate how psychological phenomena such as emotions, thoughts or behavior naturally unfold in everyday life, researchers often rely on experience sampling methodology (ESM; Hektner & Csikszentmihalyi, 2002). In a typical ESM study, participants are expected to complete a series of short surveys on their mobile device about their real-time experiences, actions and current context, while they continue with their daily routine. Although the repeated and real-life character of people's momentary responses allows for unmatched ecological insight into the temporal and situational dynamics of psychological constructs (Mestdagh & Dejonckheere, 2021), these same qualities also make ESM particularly burdensome for participants to compliantly carry out (Piasecki et al., 2007). Throughout the study course, subjects are forced to interrupt their ongoing activities numerous times, in order to mentally access, reflect upon, and rate the various internal states or behaviors under study (Scollon, Prieto, & Diener, 2009). Moreover, as most ESM studies adopt (semi-)random sampling schemes (Vachon et al., 2019; van Roekel, Keijsers, & Chung, 2019; Wrzus & Neubauer, 2022), the anticipation of unpredictable measurement prompts creates a considerable cognitive load for participants, who may feel they need to be constantly available for the next assessment (Conner & Lehman, 2012).

When participants experience an ESM protocol as intensive or demanding, they may start to deliberately skip assessments, which leads to poor response rates or even premature study termination (Morren et al., 2009; Stone et al., 2003; Vachon et al., 2019). The resulting missing data do not only undermine the statistical power to obtain stable estimates; if missingness is systematic, these estimates may also be biased (Courvoisier, Eid, & Lischetzke, 2012; Eisele et al., 2020; Sun, Rhemtulla, & Vazire, 2021). To reach valid conclusions about the real-life patterns and associations of psychological and behavioral variables, ESM research could thus benefit from procedures or strategies

☆ We made all data, materials and code (R and MATLAB) publicly available at the Open Science Framework: https://osf.io/ctznf/.

* Corresponding author. Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Leuven, 3000, Belgium.
E-mail address: egon.dejonckheere@kuleuven.be (E. Dejonckheere).

https://doi.org/10.1016/j.chb.2024.108360
Received 27 November 2023; Received in revised form 28 June 2024; Accepted 29 June 2024
Available online 2 July 2024
0747-5632/© 2024 Published by Elsevier Ltd.

that maximize response compliance (Ram et al., 2017). In this study, we examine whether instantaneously rewarding the completion of momentary ESM assessments by means of a gamification design could positively impact participants' response engagement.

1. Response compliance in experience sampling

Meta-analyses on ESM study compliance in different research domains find average response rates around 79%, and various unsystematic pooled data analyses report percentages of a similar magnitude (see Table 1 for a non-exhaustive overview). While this point estimate may be reassuring, all articles also recognize considerable between-study variation in average response rates, and even within ESM studies research reveals notable differences between participants (e.g., Messiah, Grondin, & Encrenaz, 2011; Rintala et al., 2019; Sokolovsky, Mermelstein, & Hedeker, 2014). As such, this heterogeneity underscores that there is room for improvement when it comes to advancing response engagement in participants of an ESM study.

Regarding design-related study factors that affect response compliance in ESM, review studies do not come to identical conclusions (see Table 1). For instance, while the systematic review of Morren et al. (2009) shows that lengthier momentary surveys predict lower response rates, other studies find no such association (e.g., Jones et al., 2019; Ono et al., 2019). This ambiguity is also evident in experimental research, with some longer survey manipulations leading to lower compliance rates (e.g., Eisele et al., 2020; Stone et al., 2003), while other studies find no evidence for such effects (e.g., Hasselhorn, Ottenstein, & Lischetzke, 2022). Findings are similarly inconclusive regarding the effect of fixed versus (semi-)random sampling schedules (e.g., Morren et al., 2009 vs. Vachon et al., 2019), study duration (e.g., Ottenstein & Werner, 2022 vs. Soyster et al., 2019), and assessment frequency (e.g., Vachon et al., 2019 vs. Hasselhorn et al., 2022). Nevertheless, when meaningful effects are determined, they always points towards poorer response rates for ESM studies that are more burdensome (i.e., sampling protocols that are unpredictable, longer and denser). The lack of systematic findings likely stems from low variability in certain design factors of the reviewed studies (e.g., due to homogenous samples or frequent non-reported study information; Ottenstein & Werner, 2022; Rintala et al., 2019), or because most ESM studies already balance the design parameters of their study to minimize participant burden (e.g., longer ESM studies adopt lower assessment frequencies; Dejonckheere & Erbas, 2021; Wrzus & Neubauer, 2022).

Regarding person-level characteristics that predict response compliance, meta-analytic findings are also not equivocal. Some authors report lower response rates in younger (e.g., Morren et al., 2009; Rintala et al., 2019) and male participants (e.g., Silvia et al., 2013; Vachon et al., 2019), but these relations are not always replicated in other studies (e.g., Courvoisier et al., 2012; Jones et al., 2019; Soyster et al., 2019). Non-linear age effects are equally inconclusive (Ono et al., 2019 vs. Wrzus & Neubauer, 2022). Again, no evidence is found for an opposite pattern of results (i.e., lower response rates in older and female participants). Finally, reviews that examine the overall effect of clinical status suggest that the qualification of a psychiatric diagnosis is unrelated to study compliance (e.g., Ottenstein & Werner, 2022; Wrzus & Neubauer, 2022). However, various more fine-grained analyses find lower response rates in samples that, for instance, suffer from substance abuse or psychosis, but not depression or chronic pain (e.g., Jones et al., 2019; Ono et al., 2019; Rintala et al., 2019; Vachon et al., 2019; although see Soyster et al., 2019).

Research that looked into the temporal patterns of study compliance finds that participants in typical ESM studies (i.e., around 7 days of sampling) gradually miss more assessments as the study continues (e.g., Courvoisier et al., 2012; Fuller-Tyszkiewicz et al., 2013; Ono et al., 2019; Rintala et al., 2019), which can be interpreted as a sign of increased study fatigue or loss of motivation. However, also curve-linear effects have been observed (e.g., Messiah et al., 2011; Silvia et al., 2013), with an uptick in participants' response engagement towards the end of the study. Within sampling days, different studies show that earlier assessments have a lower likelihood of being completed (e.g., Courvoisier et al., 2012; Messiah et al., 2011; Rintala et al., 2019), but this effect is partly determined by the selected starting time of the sampling schedule (e.g., starting at 9 a.m. versus noon; Silvia et al., 2013).

2. Incentives in experience sampling

Perhaps the most robust finding related to study compliance in ESM is the positive effect of (financially) compensating participants for their effort. Various systematic reviews document average response rates that are considerably higher when ESM studies offer a monetary reward for participation versus when they do not (e.g., Morren et al., 2009; Ottenstein & Werner, 2022; Wrzus & Neubauer, 2022). (Quasi-)experimental studies that looked specifically at this research question come to similar observations. Using a within-person experimental design, Ludwigs et al. (2020), for instance, found that compensating participants of an ESM study on national happiness (in the form of a €50 voucher) led to higher response rates. Similarly, Harari et al. (2017) showed that students' compliance to actively log mood-related events in a self-tracking app was higher when personalized feedback was complemented with a financial or prize reward as opposed to course credits. Finally, Lynn (2001) demonstrated that households better adhered to the completion of multiple daily diaries when a financial incentive was introduced.

In contrast, other aspects related to monetary compensation seem to play a minimal role. For instance, regarding the actual amount, most meta-analyses find that ESM studies with higher financial incentives do not necessarily report higher average response rates (Ottenstein & Werner, 2022; Wrzus & Neubauer, 2022; although see Vachon et al., 2019). This is probably because researchers already naturally adapt participation reimbursements to the anticipated study demands. Moreover, anecdotal evidence even suggests that an unusually large financial compensation (e.g., $250) can elicit adversarial effects, attracting poorly committed participants that produce high rates of missing data (Stone, Kessler, & Haythornthwaite, 1991). Second, whether ESM studies use a fixed versus incremental incentive strategy (i.e., contingent on the number of surveys that a participant actually completes) does not seem to predict differences in study compliance either (Jones et al., 2019; Mölsä et al., 2022; Ottenstein & Werner, 2022; Wen et al., 2017). Nevertheless, experimental ESM studies that isolate the effect of incremental incentive strategies on people's response rates are lacking, creating uncertainty about the true causal impact of this reward scheme (van Berkel, Ferreira, & Kostakos, 2017; van Berkel & Kostakos, 2021).

Another common way to encourage response behavior in ESM participants, one that more appeals to people's intrinsic motivation, is to provide them with (personalized) feedback (van Berkel & Kostakos, 2021). Hsieh et al. (2008), for instance, found in a between-person experiment that response rates were significantly higher when participants could inspect graphical summaries of their personal data with each completed ESM measurement. Moreover, this momentary feedback effectively mitigated the typical decline in compliance towards the end of the study. In contrast, Newcomb et al. (2018) found little difference in the response rate of participants who were randomly assigned to receive a weekly overview of their answers versus those who did not receive any individualized feedback. Similarly, in their systematic review, Ottenstein and Werner (2022) did not establish higher average response rates in studies that provided personal feedback at the end of the ESM protocol. In sum, presenting individualized feedback could be a viable incentive strategy (van Berkel et al., 2019), but the modalities (e.g., the periodicity, mode of delivery, type of visualization, etc.) that maximize its potential warrant further investigation.

3. Gamification in experience sampling

In contrast, an incentive strategy that has not received much

Table 1
A (non-exhaustive) overview of methodological studies that examined factors related to study compliance in ESM.

				Design factors				Participant factors			Temporal effects		Financial incentive effects		
Authors (year; *k* or *n*)	Domain	Average response rate	Drop-out	Survey length	Sampling scheme	Study duration	Assessment frequency	Age	Gender (male)	Clinical status	Time of study	Time of day	Yes/no	Amount	Response contingent
1. Systematic review/meta-analysis															
Jones et al. (2018; *k* = 126)	Substance use	75%		*ns* (mins)		*ns*	*ns*	*ns*	*ns*	–					*ns*
Morren et al. (2009; *k* = 62)	Pain	83%	19%	- (# items)	*ns*	*ns*	*ns*	+	*ns*				+		
Ottenstein and Werner (2022; *k* = 488)	Mental health	77%	11%	*ns* (mins)		–	*ns*	*ns*	*ns*	*ns*/–			+	*ns*	*ns*
Vachon et al. (2019; *k* = 220)	Mental health	79%	7%	*ns* (# items)	+ (fixed)	*ns*	–	*ns*	–	*ns*/–				+	
Wrzus and Neubauer (2022; *k* = 477)	Miscellaneous	79%	11%			*ns*	*ns*	*ns*	*ns*	*ns*			+	*ns*	
2. Unsystematic pooled data analysis															
Ono et al. (2019; *n* = 701)	Pain	85%		*ns* (# items)		*ns*	*ns*	+/∪	*ns*	*ns*	–				
Rintala et al. (2018; *n* = 1717)	Mental health	78%		*ns* (# items)				+	–	*ns*/–	–	+			
Silvia et al. (2013; *n* = 450)	Mental health	71%						*ns*	–	*ns*/–	-/∩	+/∪			
Soyster et al. (2019; *n* = 496)	Mental health	82%		*ns* (# items)		*ns*		*ns*	*ns*	*ns*					
3. Experimental analysis															
Eisele et al. (2020; *n* = 163)	Affect	81%	1%	- (# items)			*ns*		–		–				
Hasselhorn et al. (2022; *ns* = 313, 282)	Affect	70%, 74%	0%, 0%	*ns* (# items)			*ns*								

Note. Empty cells indicate that the effect on average response rate was not investigated, *ns* represents a non-significant (linear) relation. Letters *k* and *n* refer to number of studies and participants, respectively. Response rate = completed/total ESM surveys sent. Drop-out rate = Early-quitting during ESM protocol/total number of participants enrolled. Survey length = in number of items or in average minutes to complete. Sampling scheme = fixed versus (semi-)random. Study duration = number of sampling days. Assessment frequency = number of surveys/day. Age in years. Gender = men versus women. Clinical status = clinical group (e.g., psychosis, depression, etc.) versus healthy controls. Time of study = effect of day count within study. Time of day = effect of survey count within day. Incentive yes/no = reward (e.g., monetary, credits, therapy, etc.) versus nothing. Incentive amount = amount of financial compensation. Response contingent = amount of compensation depended on compliance or not. ∪ or ∩ = shape of the significant quadratic relation.

attention in ESM thus far, but that could be promising if adopted correctly, is gamification (Long et al., 2023). Gamification is broadly defined as *the use of game design elements in non-game contexts, often to improve user experience and engagement* (Deterding et al., 2011). Gamification research is inherently interdisciplinary, with roots in game design, behavioral economics and motivational psychology (De Croon et al., 2021). With the increasing uptake of mobile technology in daily life, gamification may serve as an excellent candidate mechanism to incentivize participants in an ESM study. Since this strategy permits researchers to include typical game elements in their design, such as points, badges, trophies, rewards, progress bars, challenges, avatars, leaderboards, etc. (Werbach & Hunter, 2012), it could make a non-game, low-stake, high-burden ESM context more interesting for individuals to engage with.

To our knowledge, van Berkel, Goncalves and colleagues (2017) are the only researchers thus far who explicitly explored the effect of gamification on people's response behavior in ESM. In a somewhat atypical ESM study, they invited 24 university students to play a mobile game (*with a purpose*; von Ahn & Dabbish, 2008) in daily life. Participants received a combination of time- and location-contingent notifications (based on mobile GPS sensing), and were asked multiple times per day to provide a relevant one-word description about their current environment, as well as to evaluate the relevance of responses submitted by other individuals that had visited that location. They could also launch the app at their own accord or play multiple game rounds per notification. In the control condition, participants played the game just like that. In the experimental condition, however, various persuasive game elements were added, such as the possibility to receive points for each submitted response (based on the crowdsourced evaluations of other players), response time restrictions to make the evaluations more challenging, and a leaderboard to compare one's accumulated score to those of other participants. Comparisons revealed that individuals in the gamified ESM protocol had slightly higher time- and location-contingent response rates, and that they provided significantly more spontaneous responses (i.e., self-initiated logs or replays after a notification). Moreover, other players judged the submissions of people in the gamified experimental condition to be of higher quality (i.e., more descriptive of one's current location), as determined via a majority voting scheme. Taken together, this study provides initial evidence for a positive effect of gamification on participants' response engagement in ESM, both in terms of quantity and quality (van Berkel et al., 2017). However, whether and to what extent this effect also manifests in more traditional ESM protocols, where participants repeatedly rate the intensity of internal experiences (e.g., emotions or symptoms as opposed to "verifiable" context labels), following sampling schemes that are entirely notification-based (e.g., semi-random assessments that cannot be self-initiated), remains uncertain.

Despite methodological research still being scarce, using game-based incentives to promote response behavior in ESM already has some recognizable benefits. From a practical perspective, this incentive strategy may come without any real-life monetary costs. Consequently, virtual rewards or compensations could be a viable alternative to motivate participants when financial reimbursements are difficult (e.g., due to limited study budget) or inappropriate (e.g., ESM in the context of therapeutic case conceptualizations; von Klipstein et al., 2020). Furthermore, when a particular research question requires a sampling scheme that is inherently intensive, gamification techniques could be used to complement other implemented incentive strategies. Another major advantage of game-based rewards is that they can be automatically awarded *in the moment*. From a theoretical perspective, the *immediacy* principle in operant conditioning states that immediate reinforcements induce faster response acquisition compared to delayed reward schemes (Miltenberger, 2015). This suggests that awarding digital micro-incentives after each completed assessment could be more effective for participants' response compliance than, for instance, a one-time delayed financial compensation at the end of the study (i.e., bulk payment; Musthag, Raij, Ganesan, Kumar, & Shiffman, 2011; van Berkel, Ferreira, & Kostakos, 2017). In that vein, the momentary reward scheme that follows from a gamification procedure is the only incentive strategy that can adequately counter non-response at the within-person level (e.g., mitigate temporal or situational effects; e.g., Rintala et al., 2020; Silvia et al., 2013; van Berkel et al., 2020), which accounts for more than 50% of the variance in survey completion (Ono et al., 2019). At the between-person level (which explains around 25% of the variance), game-based (extrinsic) incentives like points or badges may be particularly successful in boosting the response rates of younger and male participants (e.g., Bittner & Shipper, 2014; Oyibo, Orji, & Vassileva, 2017), two characteristics that have been related to poor study compliance in ESM (e.g., Rintala et al., 2019).

4. The current study

To examine the effect of a gamified momentary reward system on people's response compliance in ESM, we set up a between-subjects experiment with two conditions. The idea was that using real-time game-based incentives to promote survey completion would motivate participants to engage with this behavior more frequently, and that it could mitigate the inevitable burden related to this intensive type of data collection. Participants in the control condition took part in a traditional ESM protocol, where they repeatedly rated multiple momentary subjective experiences (emotions, life domain evaluations) in everyday life. Participants in the experimental condition were presented with the exact same task, but after each completed survey they immediately received a small sum of virtual coins that they could save or spend in exchange for various extrinsic (fun facts) or intrinsic (personalized graphs) in-app incentives. In this gamified ESM protocol, we included different persuasive game-based elements to stimulate participants to complete as many assessments as possible (see Table 2 for a summary).

Although we primarily designed our momentary gamification

Table 2
The persuasive architecture of our gamified ESM protocol.

Game design element	Operationalization	Persuasive mechanism
1. Coins	Points earned with each completed assessment	Reward: Immediate positive reinforcement, token strategy
2. Streak	Additional points earned with an uninterrupted series of completed assessments	Challenge: Maintain streak across assessment (or negative reinforcement: Avoid streak interruptions)
3. Fireworks	Supportive animation visualized after each completed assessment	Fun/Playfulness: Visual and aesthetically pleasing feature to highlight desired behavior
4. Coin wallet	Real-time information about the amount of collected coins	Progress feedback: Communication about coins currently accumulated
5. Coin shop	Virtual store to purchase rewards with collected coins	Autonomy: Freedom to spend collected coins to any available reward, at any time
6. Fun facts	Assortment of curious facts about different domains to purchase	Reward: Virtual reward to incite extrinsic motivation
7. Feedback graphs	Assortment of different (updating) personal data visualizations to purchase	Reward: Virtual reward to incite intrinsic motivation
8. Banner	Real-time information about collected rewards in home screen	Performance feedback: Trophy cabinet that displays purchased rewards

procedure to promote survey completion, we also explored its collateral effects on other crucial response-related aspects. After all, in order to collect ecologically valid data in ESM, it is not only of critical importance to ensure that non-response is limited and unsystematic, participants' completed momentary evaluations should also be free of bias and reliable (Ram et al., 2017). In terms of bias, the latency of people responding to an ESM notification has received some attention (e.g., Delespaul, 1995; Eisele et al., 2021), with longer delays suggesting a higher degree of self-selected sampling moments (Christensen et al., 2003). When participants consistently postpone the completion of momentary surveys to more convenient times (e.g., when their negative emotions are less intense), the assumed semi-random assessments may no longer represent the true distribution of particular events in a person's everyday life (e.g., emotionally intense events are underrepresented). We therefore explored if gamifying survey completion also resulted in shorter response latencies. Next, in terms of reliability, a sizeable discrepancy between two identical items in the same momentary survey could be an indication of careless responding (Dejonckheere et al., 2022), as do very fast completion times (Jaso, Kraus, & Heller, 2021; Schneider, May, & Stone, 2018). While it is possible that momentary game-based rewards motivate participants to interact with upcoming ESM questionnaires, this incentive strategy does not necessarily imply that they also complete that survey with sufficient diligence or attention. In fact, some scholars have argued that gamification could shift the focus away from the actual task, undermining the reliability of participants' responses (e.g., Keusch & Zhang, 2017). Consequently, we investigated to what extent our gamification procedure impacted these quality indicators of people's submitted responses.

5. Materials and methods

5.1. Openness and transparency

The current experiment was ethically approved by the Social and Societal Ethics Committee of KU Leuven (*Exploring the effect of gamification strategies on participants' response rates and reliability in experience sampling studies.*; G-2021-4628-R3). We preregistered our study design, research questions and statistical analyses at the AsPredicted platform (#101798), and describe exploratory analyses as such. To reproduce our findings, readers can download the original data and analysis code (in MATLAB and R) from OSF (https://osf.io/ctznf/) or they can freely use the baseline and ESM surveys created for this study from the public m-Path library (www.m-Path.io: *Gamification Study – Replication (baseline/ESM)*) for direct replication purposes. These questionnaires include the installation of our coin store, the response streak computations, and our coin award strategy.

We confirm that our sample size determination is explained, and that we report on all data exclusions and manipulations. We collected a battery of person-level baseline measures to address some substantive research questions (i.e., related to people's actual ESM ratings) that are beyond the scope of the current report, but we provide a full survey list in Supplemental Materials 1.

5.2. Power calculation and participant recruitment

To establish the required sample size for our experiment, we conducted a power analysis prior to the collection of the data. Based on the modest meta-analytic effect sizes of gamification on behavior in related disciplines (e.g., Ba et al., 2020; Mazeas et al., 2022; Sailer & Homner, 2020), we conservatively expected our real-time gamified ESM manipulation to have a small to moderate impact on people's response rate. Assuming an average probability (p) difference of 0.10 in response engagement between the gamification versus control condition, we randomly sampled simulated participants from two normal distributions, with $p_{\text{gamification}} \sim N(0.65,0.20)$ and $p_{\text{control}} \sim N(0.55,0.20)$. From this simulated sample, we fitted a two-level random intercepts model for time series with 100 momentary surveys at the within-person level and our gamification manipulation as a dummy predictor at the between-person level. Our simulation took into account that we aimed to allocate participants unevenly to the experimental (2/3) versus control condition (1/3), which had to enable analyses into the continuous effects of gamification exposure on participants' response motivation in the experimental condition. We repeated this process 200 times and evaluated the number of times we secured significant differences in response compliance between our experimental and control group (with $\alpha = .05$). To achieve a statistical power of 90%, this implied the recruitment of at least 190 participants.

To obtain our predetermined sample size, we enrolled 200 Dutch-speaking adults in our study, accounting for some drop-out. We had to exclude seven participants from our final analyses due to phone malfunctioning ($n = 2$), erroneous sampling schemes ($n = 3$), early study termination ($n = 1$), or data withdrawal for unknown reasons ($n = 1$). In the end, this lead to a final sample size of 193 participants (81.3% female) with an average age of 22.67 ($SD = 8.52$). 65 individuals were randomly assigned to the control condition (78.5% female; age $M = 22.83$, $SD = 8.31$), while 128 people received the gamification manipulation in the experimental group (82.8% female; age $M = 22.59$, $SD = 8.66$). No meaningful age ($t = .18$, $d = .03$, $p = .851$) or gender ($X^2 = .52$, $p = .470$) differences were observed between conditions.

Our total sample consisted of a mix of students who took part in our study in exchange for 1.5 course credits ($n = 108$) and volunteers from the general community who were paid €15 ($n = 85$). Both types of participants were evenly distributed across conditions ($X^2 = .65$, $p = .420$). Compared to other ESM studies that were conducted at KU Leuven and that had a similar participation burden (e.g., Carlier et al., 2022; Dejonckheere et al., 2018; Kalokerinos et al., 2022), these reimbursements were relatively low and not contingent upon participants' actual compliance rates. This had to ensure that participants' response behavior was more a function of gamification elements during the ESM protocol (i.e., virtual coins that could be spent on fun facts and/or personal data visualizations), and less by real-world incentives (i.e., credits or money). In this way, we aimed to maximize the potential effect of our momentary gamification manipulation. All participants provided informed consent.

5.3. Procedure

Our ESM study took place between July and October 2022. To prevent that COVID-19 restrictions compromised data collection, we decided to meet participants online in small groups via video call. During this virtual session, we broadly introduced our study as an investigation of the dynamic interplay between people's momentary emotions and important life domains in everyday life (which is, in fact, another actual research aim of this data collection). When describing the general study procedure, we explicitly left our research interest in the effect of real-time gamification on people's response behavior, in order not to instigate potential demand effects. Next, participants downloaded our ESM app m-Path on their smartphone (Mestdagh et al., 2022), and we asked them to give electronic informed consent. Our introduction meeting ended with the completion of a battery of baseline questionnaires, irrelevant for this report (see Supplemental Materials 1 for a complete overview).

5.3.1. ESM protocol

The actual ESM phase began the next day. For 10 consecutive days, participants were instructed to answer to 10 m-Path notifications per day (i.e., 100 ESM notifications in total), which we scheduled semi-randomly between 9 a.m. and 7 p.m. This sampling scheme enabled a representative picture of participants' day-time activities (as opposed to purely random schedules), while natural behavior was not jeopardized as a result of predictable measurements (as opposed to fixed schedules; Dejonckheere & Erbas, 2021). The average time window between two

subsequent notifications was 60 min (*SD* = 25 min), and participants were instructed to complete the momentary assessment immediately upon receiving it. If they did not attend to a notification within 10 min, the survey expired and the assessment was missed.

At each measurement occasion, we invited participants to rate the intensity of six positive and six negative emotions (from *Not at all* to *Very much*), as well as the momentary evaluation of seven life domains (from *Very positive* to *Very negative*), on a 10-point slider scale. One randomly selected emotion item was presented twice to evaluate the momentary reliability of people's responses (Dejonckheere et al., 2022). A multiple choice question about the most important life domain(s) at that time concluded the ESM questionnaire (see Supplemental Materials 2 for a full list of ESM items).

5.3.2. Gamification manipulation

Using a between-person design, we randomly assigned participants to either the experimental or control condition (ratio 2:1). The ESM protocol in the experimental condition differed from the one in the control group in that we immediately rewarded the completion of every scheduled momentary survey with a remuneration of 10 virtual coins (including fireworks). To further incentivize uninterrupted completion streaks, we complemented this fixed allowance scheme with a variable bonus that increased with 2 coins per completed subsequent assessment. Not only did this introduce an element of challenge, this practice also had the potential to counter the fact that people often miss ESM assessments consecutively (Rintala et al., 2020). The procedure of this reward-based behavioral economy (Hursh, 1984) was not verbally explained, but participants gradually experienced it while completing ESM measurements (see Panel A in Fig. 1). Across assessments, the collection of these virtual coins functioned as a token strategy (Kazdin, 1977), where participants could save earned coins in a personal wallet, and spend them in an in-app coin store any time (see Panel B in Fig. 1).

In this store, which we developed specifically for this study, an initial sum of 20 coins could be exchanged for a curious fun fact about one of 10 domains of interest (e.g., History – *In 1807, Napoleon Bonaparte had to 'battle' against a large colony of rabbits.*). This type of reward was used to appeal to people's (short-term) extrinsic motivation (i.e., gratification is unrelated to the reinforced behavior; Harlen & Deakin Crick, 2003). The list of facts was self-compiled (see Supplemental Materials 3 for a complete fun fact overview), but we ensured that they all contained an element of surprise or thought incongruence, as traditional reinforcement theories consider this critical to motivate learning behavior (e.g., Rescorla, 1972). There were five fun facts per domain, and with each purchase one fun fact was presented (see Panel C in Fig. 1). With every fun fact bought, the price of the other fun facts within that domain increased with 10 coins to roughly maintain a similar level of challenge. A banner in the app's home screen displayed all unlocked facts, to repeatedly remind participants about their collected rewards (e.g., trophy cabinet; Robson et al., 2015).

Alternatively, participants could spend 60 coins on various (continuously updating) graphical visualizations of their personal data (e.g., time series of their *cheerful* ratings, pie chart of their domain importance, etc.; see Panel C in Fig. 1). This type of reward had to address participants' (long-term) intrinsic motivation (i.e., purposefully engaging with the assessments to receive feedback and improve self-insight; Buckley & Doyle, 2016). The graphs were slightly more expensive than the fun facts because they required sufficient completed data points to have their rewarding effect. Again, all purchased data visualizations could be (re) accessed at people's own convenience in the app's home screen (see Panel D in Fig. 1).

5.3.3. Manipulation check

After the experiment was finished, participants in the gamification condition ($n = 128$) completed a final 6-item ESM survey to provide us with a better understanding of the persuasive mechanism of our reward procedure. On a 7-point slider scale that ranged from *Not at all* to *Very much*, we evaluated to what extent they (1) thought the coin store was user friendly, (2) wanted to respond to the surveys to earn coins, (3) tried to maintain a response streak, (4) wished to spend coins on fun facts, (5) wished to spend coins on personal graphs, (6) found these incentives rewarding.

5.4. Statistical analyses

On the criterion side of our models, we examined the effect of our outlined real-time gamification strategy on four outcomes of interest. Regarding actual response engagement, we investigated (a) whether people in the experimental condition, on average, completed more ESM assessments than participants in the control group (i.e., *response compliance*; preregistered), and (b) whether they generally responded faster to the measurement notifications (i.e., *response latency*; exploratory). Participants' response compliance was operationalized as a binary variable that indicated whether they had interacted with the scheduled ESM notification or not (e.g., Hasselhorn et al., 2022), while their response latency represented the time interval between receiving an ESM notification and opening the associated survey (e.g., Eisele et al., 2021). Regarding the quality of their provided responses, we analyzed (c) the inconsistency of identical ratings (i.e., *response unreliability*; preregistered[1]), as well as (d) the speed with which they completed the momentary survey (i.e., *response completion time*; exploratory). Participants' response unreliability was derived from the discrepancy between the ratings from the randomly selected test-retest emotion pair at each momentary assessment (e.g., Dejonckheere et al., 2022). Response completion time referred to the (log-transformed) time that elapsed between opening the ESM survey and submitting the survey responses (e.g., Jaso et al., 2021). We considered a logarithmic variant to minimize the effect of very long completion times (i.e., skew normalization; Feng et al., 2014).

On the predictor side of our models, we included a fixed set of variables in a two-level mixed effect model to account for the nested structure of the data (i.e., different ESM notifications per individual).[2] For all outcomes we used *linear* regression models, except for *response compliance* (a binary outcome), for which a *logistic* model was fitted (see Equation (1) for an example). At the notification-level, we included a person-specific intercept β_{0i} and two time-related predictors to account for the established temporal trends in participants' response behavior (e.g., Ono et al., 2019; Rintala et al., 2020; exploratory). *Time of study* indicated the exact moment t a notification was sent for individual i during the entire ESM study in continuous days from 0 to 9, with zero indicating the first day of the protocol. *Time of day* included the exact same information, but at a daily level. This variable represented the continuous hours within a day and ranged from 0 to 10, with zero indicating 9 a.m. (the start-point of each sampling day). For each predictor, temporal slopes β_{1i} and β_{2i} were allowed to vary across participants and residual scores ε_{ti} were assumed to be normally distributed with $\varepsilon_{ti} \sim N(0,\delta)$. Next, at the intercept's second level, we included three person-level predictors. First, *categorical gamification* indicated whether participant i received the experimental manipulation or not (0 = control, 1 = gamification; preregistered). Second, we included *continuous*

[1] Because prior work on the impact of gamification on the quality of people's responses was inconclusive (e.g., van Berkel, Ferreira, & Kostakos, 2017 vs. Keusch & Zhang, 2017), we agnostically predicted that there would be no differences between conditions in terms of response unreliability.

[2] Note that, in our preregistration, we originally proposed to use ANOVA and simple linear regression to evaluate group differences in ESM compliance. However, because previous work established robust within-person dynamics of response engagement over time (i.e., within-day and -study effects; Rintala et al., 2018), we deemed the inclusion of these temporal covariates essential. This could only be achieved in a multilevel framework, where within-person processes and between-person differences are effectively separated.

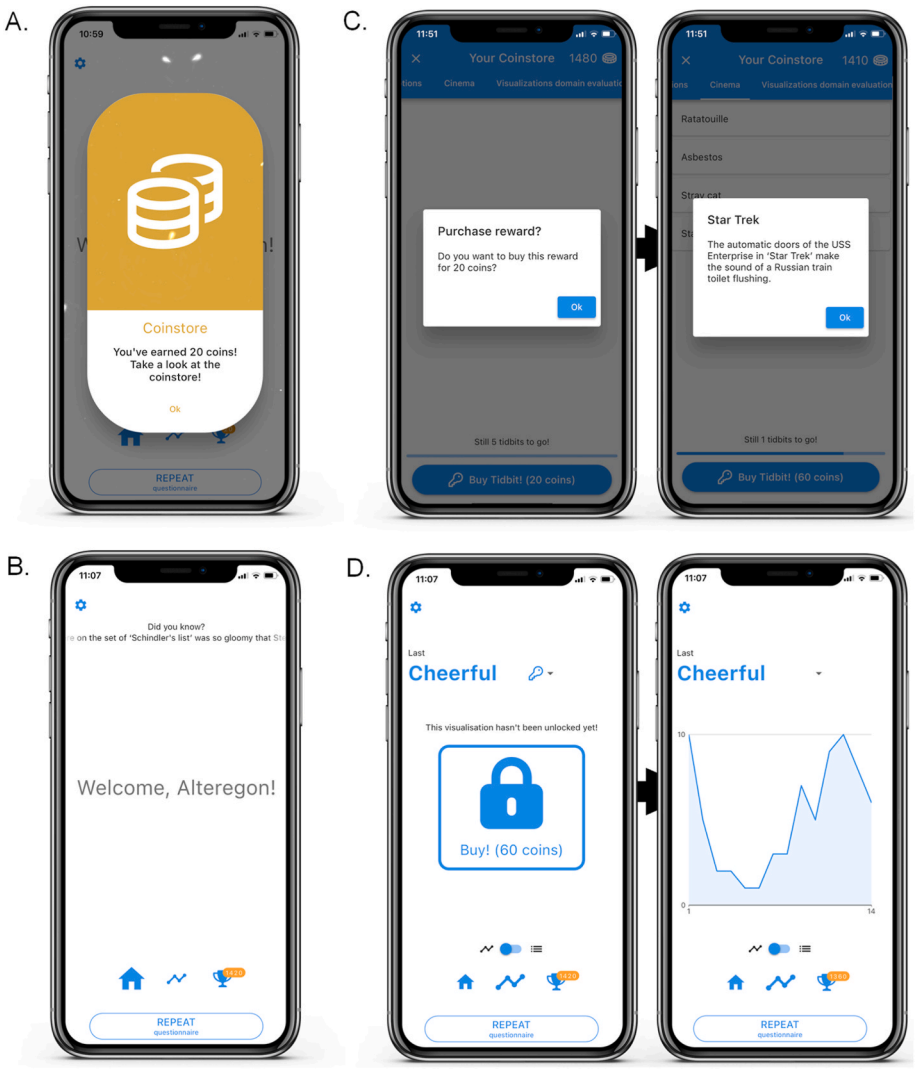

Fig. 1. Different m-Path screenshots of our in-app gamification procedure in the experimental condition. Panel A shows the reward message presented immediately after survey completion about the earned coins. Panel B shows the app's home screen with the fun fact banner, the coin store access, and amount of available coins. Panel C and D illustrate the purchase process of fun facts and personal data visualizations, respectively.

gamification, which represented the extent with which participant *i* in the experimental condition actually engaged with our gamification procedure (preregistered). Specifically, this predictor reflected the percentage of collected coins that a participant had spent in the coin store, with 0 indicating no purchases were made and a participant thus not received any virtual rewards (for individuals in the control condition this variable was 0 by default). Third, based on the observed individual differences in ESM response compliance in previous studies (e.g., Rintala et al., 2019; Silvia et al., 2013), we also controlled for *gender* (0 = female, 1 = male; exploratory).[3] We did not consider age differences due to the restricted age variability in our sample (i.e., 75% of participants had an age between 17 and 23). Together, all γ's represented the model's fixed effects, while all η_i's reflected its random effects. The Level 2 variables predict differences in the person-specific intercepts (i.e., the individual level of the response variable at the first time point of the study). Because they only predict the model's intercept and not the temporal slopes, their effect also represents how a particular response variable differs *on average* between conditions.

Level 1:

$$Response\ compliance_{ti} = \beta_{0i} + \beta_{1i} Time\ of\ study_{ti} + \beta_{2i} Time\ of\ day_{ti} + \varepsilon_{ti} \quad (1)$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} Categorical\ gamification_i + \gamma_{02} Continuous\ gamification_i + \gamma_{03} Gender_i + \eta_{0i}$$

$$\beta_{1i} = \gamma_{10} + \eta_{1i}$$

$$\beta_{2i} = \gamma_{20} + \eta_{2i}$$

For these main analyses, we ran all models with and without potential outliers, based on Mahalanobis distance calculation (preregistered). The outlier-free models are presented in Supplemental Materials 5, but their estimates did not largely differ from the full-sample ones we report in the main article.

In a series of secondary analyses, we aimed to further unpack the persuasive mechanisms of our coin-based momentary gamification

[3] In Supplemental Materials 4, we also explored gamification × gender interactions, since previous work showed that male participants may be more susceptible to game-based reward strategies (e.g., Oyibo et al., 2017). In the current study on response compliance in ESM, we found little evidence for this idea. However, it should be noted that our original sample size was not determined to test for these interactions with sufficient statistical power.

procedure (exploratory). First, we examined each participant's longest response streak and evaluated differences between conditions via an independent samples *t*-test. This allowed us to establish whether the added coin bonus incited participants to maintain a series of uninterrupted responses. Second, we explored what items participants in the gamification condition actually bought, to see whether differences emerged between our extrinsic (fun facts) and intrinsic (personal graphs) rewards. Finally, we related participants' answers from the manipulation check to their outcome measures, to see how their perception of the various incentive strategies correlated with their actual response behavior.

6. Results

Descriptive statistics and correlations are presented in Table 3. The average compliance in our entire sample was 58.31% (SD = 21.13%, range = 1%–96%), which is considerably lower than the meta-analytic estimates reported in Table 1. This may be explained by the fact that we provided comparatively lower financial reimbursement fees (in relation to other studies at our department), that our compensations did not depend on people's actual compliance rates (which led some participants to only interact with a single assessment), and that we did not impose an a priori compliance cut-off for our final analyses (which is often the case in substantive ESM studies). One participant in the control condition prematurely dropped out from the study for unknown reasons (i.e., other than phone malfunctioning).

6.1. The effect of instantly gamifying response engagement

Table 4 presents the model results for each of the four investigated response outcomes. First, for compliance, at the notification-level, we determined a within-study effect, with participants' probability to respond to a momentary measurement prompt significantly decreasing as the ESM study continued (i.e., a likelihood[4] difference of 20% between the first versus last day of sampling when we control for the effect of other predictors[5]). In contrast, a within-day effect was not significant, suggesting that participants' probability to respond did not meaningfully change throughout the day (i.e., a likelihood difference of 3% between 9 a.m. versus 7 p.m.). At the person-level, contrary to our expectations, people in the gamification condition, on average, did not automatically have higher response rates (i.e., a difference of 1% in likelihood to respond when being in the control versus gamification condition; see Panel A in Fig. 2). However, when additionally taking into account to what extent participants in the gamification condition actually engaged with our reward procedure, we did observe a significant and positive effect (however, this effect became non-significant in the outlier-free models with p = .052; see Supplemental Materials 4). This is because a considerable number of people in this condition did not actually interact with our coin store (n = 45; see Panel A in Fig. 3). In contrast, participants who did frequently buy in-app incentives with their earned coins had higher response rates (e.g., an increase of 9% in likelihood to respond when all versus none of the collected coins were spent). This could suggest that rewarding people with virtual coins in itself did not promote the desired behavior of increasingly responding to ESM notifications, but rather did the opportunity to exchange these coins for in-app rewards like fun facts or personalized graphs (see also Panel B in Fig. 3). Finally, male participants, on average, had

Table 3
Descriptive statistics and correlations between all measures.

	Mean	$SD_{between}$	SD_{within}	1	2	3	4	5	6	7	8	9	10	11
Outcome measures														
1. Response compliance (%)	58.31	21.13		–										
2. Response latency (sec)	114.30	46.79	142.18	−0.28***	–									
3. Response unreliability	0.55	0.31	0.91	−0.09	0.001	–								
4. Response completion time (sec)	63.58	19.39	18.38	−0.30***	0.24***	−0.21**	–							
5. Longest response streak	11.80	8.68		0.76***	−0.29***	−0.07	−0.20**	–						
6. Percentage coins spent	27.05	39.52		0.16*	−0.18*	−0.06	0.06	0.16*	–					
Persuasive game elements (manipulation check)														
7. Found coin store user friendly	4.17	2.51		0.22*	−0.08	0.06	0.02	0.09	0.15					
8. Respond to earn coins	2.05	1.87		0.11	−0.20*	0.05	−0.03	0.09	0.18*	0.57***				
9. Desire to maintain response streak	3.30	2.31		0.34***	−0.07	0.09	−0.11	0.28**	0.03	0.72***	0.52***			
10. Desire to spend coins on facts	2.55	2.28		0.22*	−0.14	−0.04	−0.13	0.22*	0.46***	0.55***	0.54***	0.48***		
11. Desire to spend coins on graphs	2.24	2.15		0.18*	−0.19*	0.12	0.01	0.16	0.42***	0.54***	0.52***	0.44***	0.59***	
12. Found incentives rewarding	3.09	2.32		0.16	−0.10	0.05	0.07	0.08	0.18*	0.72***	0.73***	0.53***	0.57***	0.55***

Note. The mean (and SD) for *percentage coins spent* in the gamification condition only is 40.79% (42.39%). Between-person correlations are presented below the diagonal. Those that involve the manipulation check items are based on n = 128, because these questions were only presented in the experimental condition. All other correlations are based on the entire sample (n = 193). For (continuous) momentary measures, the (average) within-person correlations are shown above the diagonal. *$p < 0.050$, **$p < 0.010$, ***$p < 0.001$.

[4] We converted logistic regression coefficients (β) into probabilities (p), with $= \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$.

[5] We always report effect sizes for the hypothetical instance where the value of (other) predictors is zero, which describes a measurement during the first day of sampling (*time of study*), around 9 a.m. (*time of day*), for female (*gender*) participants in the control condition (*categorical gamification*), who did not spend any virtual coins in the coin store (*continuous gamification*).

Table 4
Summary of mixed effects models for each outcome.

		Response Compliance				Response Latency				Response Unreliability				Response Completion Time (Log)			
Predictors	**Values/Levels**	Estimate	*t*	*p*	ES	Estimate	*t*	*p*	ES	Estimate	*t*	*p*	ES	Estimate	*t*	*p*	ES
1. Notification-level																	
Intercept		0.78	6.21	<0.001		109.87	15.38	<0.001		0.51	11.67	<0.001		4.32	130.14	<0.001	
Time of study	0 (day 1) – 9 (day 10)	−0.09	−11.08	<0.001	−20%	2.20	4.22	<0.001	20 s	−0.02	−4.73	<0.001	−0.18	−0.04	−16.72	<0.001	−23 s
Time of day	0 (9 a.m.) – 10 (7 p.m.)	0.01	1.60	0.110	3%	−1.08	−2.21	0.027	−11 s	0.01	1.91	0.056	0.1	−0.01	−4.92	<0.001	−7 s
2. Person-level																	
Categorical gamification	0 (control) – 1 (gamification condition)	−0.06	−0.39	0.695	−1%	4.65	0.58	0.559	5 s	0.20	3.91	<0.001	0.20	−0.04	−0.82	0.412	3 s
Continuous gamification	0 (0%) – 1 (100% collected coins spent)	0.44	2.19	0.028	9%	−22.93	−2.71	0.007	−23 s	−0.17	−2.95	0.003	−0.17	0.08	1.20	0.232	6 s
Gender	0 (female) – 1 (male)	−0.58	−2.90	0.004	−14%	4.35	−1.17	0.665	4 s	−0.06	0.45	0.241	−0.06	0.01	0.21	0.832	1 s

Note. Effect sizes (ESs) indicate the difference between the two most extreme values (or levels) of a continuous (or categorical) predictor, when we control for the effect of other predictors.

compliance rates that were significantly lower (i.e., a decrease in likelihood to respond of 14% for male versus female subjects).

Second, regarding response latency, we observed significant within-study and within-day effects. Participants generally took longer to open an ESM notification towards the end of the study (e.g., an increase of 20 s at the end versus begin of the ESM protocol), as well as during morning hours (e.g., an increase of 11 s at 9 a.m. versus 7 p.m.). Similar to people's response compliance, merely being exposed to a gamification procedure did not produce shorter latencies to open a momentary ESM notification (e.g., a difference of 5 s between control versus gamification condition; see also Panel B in Fig. 2). Rather, actively engaging with this reward scheme and exchanging coins for virtual rewards was associated with shorter response latencies (e.g., people who spent all their coins opened the ESM survey 23 s faster than those who spent none). An effect of gender was not significant.[6]

Third, with respect to response unreliability, we found that people's momentary test-retest discrepancies diminished towards the end of the study (i.e., a difference of 0.18 rating points on the 10-point slider scale on the first versus last day of sampling). In contrast, the temporal trend within a day was not significant. Regarding our experimental manipulation, we initially observed how gamifying response engagement somewhat paradoxically undermined the reliability of people's responses (see also Panel C in Fig. 2). Participants who received virtual coins for each completed ESM measurement showed significantly larger discrepancies between their momentary test and retest ratings (i.e., a small difference of 0.20 rating points between experimental and control condition). However, this adverse effect was particularly prevalent in people who did not actively exchange their coins for virtual incentives (i.e., they had an average test-retest discrepancy of 0.72; see also Panel C in Fig. 3). The more coins a participant spent, the more his or her response reliability approximated that one of people in the control condition (i.e., an estimated average test-retest discrepancy of 0.51 for people in the control condition versus 0.55 for people in the experimental condition who spent all of their coins). Finally, a gender-effect was not significant.

Fourth, for completion times, we observed significant temporal trends across and within study days. Towards the end of the ESM study and the end of each day, participants generally completed the ESM surveys faster (i.e., a reduction of 23 and 7 s, respectively). Instantly receiving virtual coins for completion, however, did not affect the speed with which participants completed the momentary measurements, nor did the possibility to spend these coins on fun facts or rewards. Gender differences were also not significant (however, in the outlier-free models male participants completed surveys faster; see Supplemental Materials 5).

6.2. *Exploring the persuasive elements of instantly gamifying response engagement*

First, we investigated whether the addition of a coin bonus encouraged people not to miss or skip momentary assessments, by evaluating whether participants in the gamification condition, on average, maintained longer uninterrupted response streaks. An independent samples *t*-test revealed that this was not the case ($M_{\text{control}} = 10.63$ versus $M_{\text{gamification}} = 12.39$; $t = -1.33$, $d = -0.20$, $p = .184$).

Second, we examined purchase behavior. On average, people in the gamification condition earned 1025 virtual coins throughout the study course ($SD = 653$ coins, median = 887, range = 30–3200 coins), but only spent 493 ($SD = 676$ coins, median = 180, range = 0–3200 coins). On average, participants spent more coins on fun facts ($M = 280$; 95% CI = [202–368]) than on personal graphs ($M = 198$; 95% CI =

[6] When we additionally included gamification × gender interactions, we observed how male participants in the gamification ESM condition actually showed *longer* response latencies (see Supplemental Materials 4 for full results).

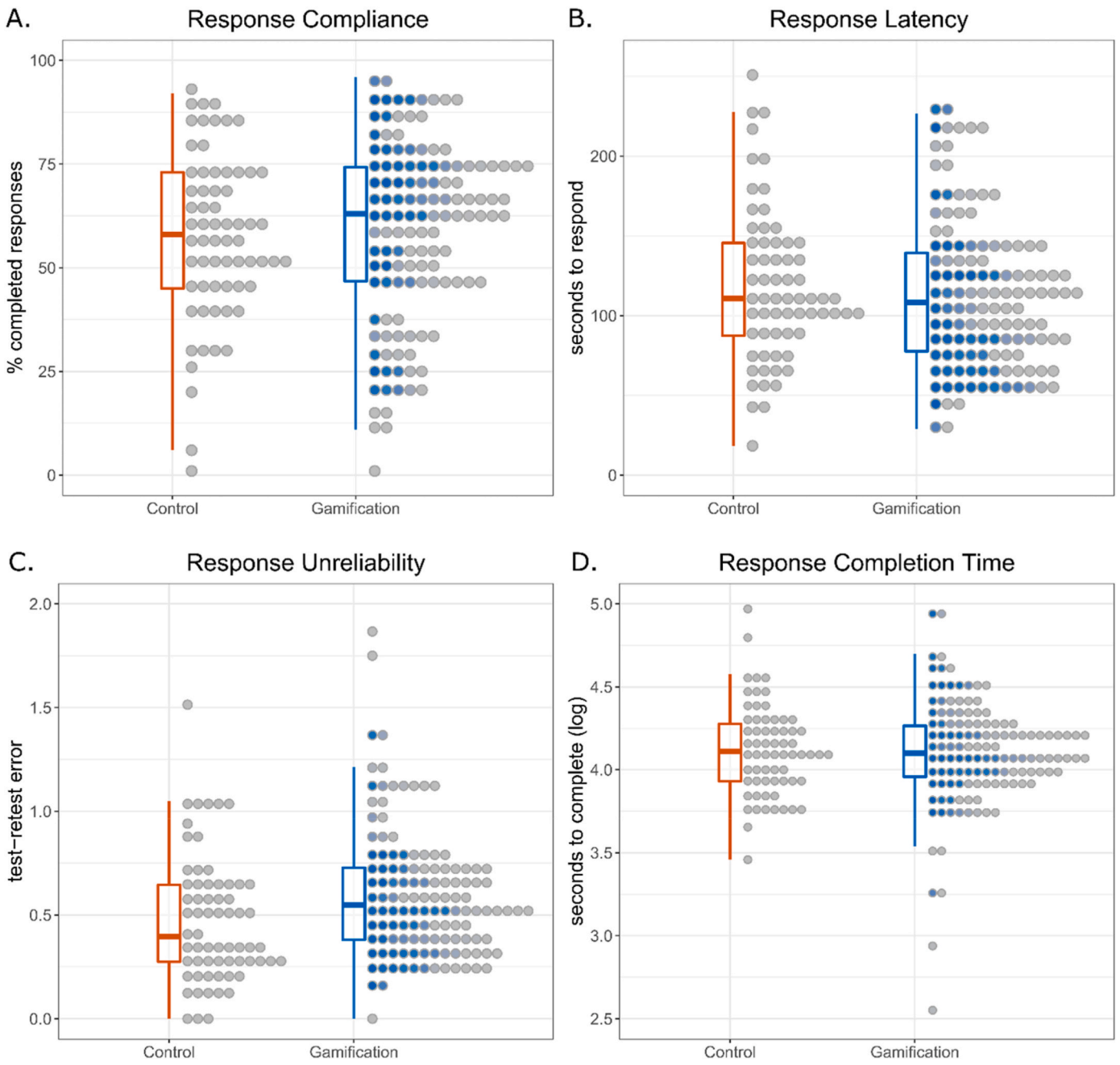

Fig. 2. Sample distribution and boxplot of the control (*red*) and gamification (*blue*) condition for each response outcome. In each panel, a dot represents the average value of an outcome for an individual. The blue gradient used to fill each case indicates to what extent that individual actually interacted with the gamification procedure, operationalized as the percentage of collected coins spent in the virtual store to buy in-app incentives (*light gray* = no purchases made; *dark blue* = many purchases made).

[140–258]), but a bootstrap procedure with 2000 resamples indicated that this difference was not significant, as the 95% confidence intervals overlapped. This suggests that intrinsic and extrinsic rewards were equally favored in terms of spending behavior.

Finally, we interpreted the results from our manipulation check (see Table 3 for a full overview of descriptives and correlations). While participants in the gamification condition found our coin store modestly user friendly (M of 4.17 on a 7-point Likert scale, $SD = 2.51$), they indicated that the virtual coins themselves did not really encourage them to respond to the momentary surveys ($M = 2.05$, $SD = 1.87$). However, the self-reported motivation to earn coins was linked to faster interactions with the notification signal ($r = -0.20$, $p = .025$), and a higher percentage of actual coins spent ($r = .18$, $p = .039$). There was no significant relation with people's response rates ($r = .11$, $p = 0.201$). Next, the challenge to maintain a response streak was found to be significantly more motivating to complete surveys ($t = 6.81$, $d = .41$, $p < 0.001$), but still only modestly so ($M = 3.30$, $SD = 2.31$). People's judgement corresponded with their actual longest completed response streak ($r = .28$, $p = .002$), and also related to higher response rates ($r = .34$, $p < 0.001$). Regarding the in-app rewards, participants did not express a strong wish to spend collected coins on fun facts ($M = 2.55$, $SD = 2.28$) or personal graphs ($M = 2.24$, $SD = 2.15$). A difference in reported motivation for these two types of incentives was not significant ($t = 1.76$, $d = -0.33$, $p = .081$), and both desires related to higher percentages of coins spent (r's > 0.42, p's < 0.001), and higher compliance rates (r's > 0.19, $p < 0.042$). Overall, participants found the incentives

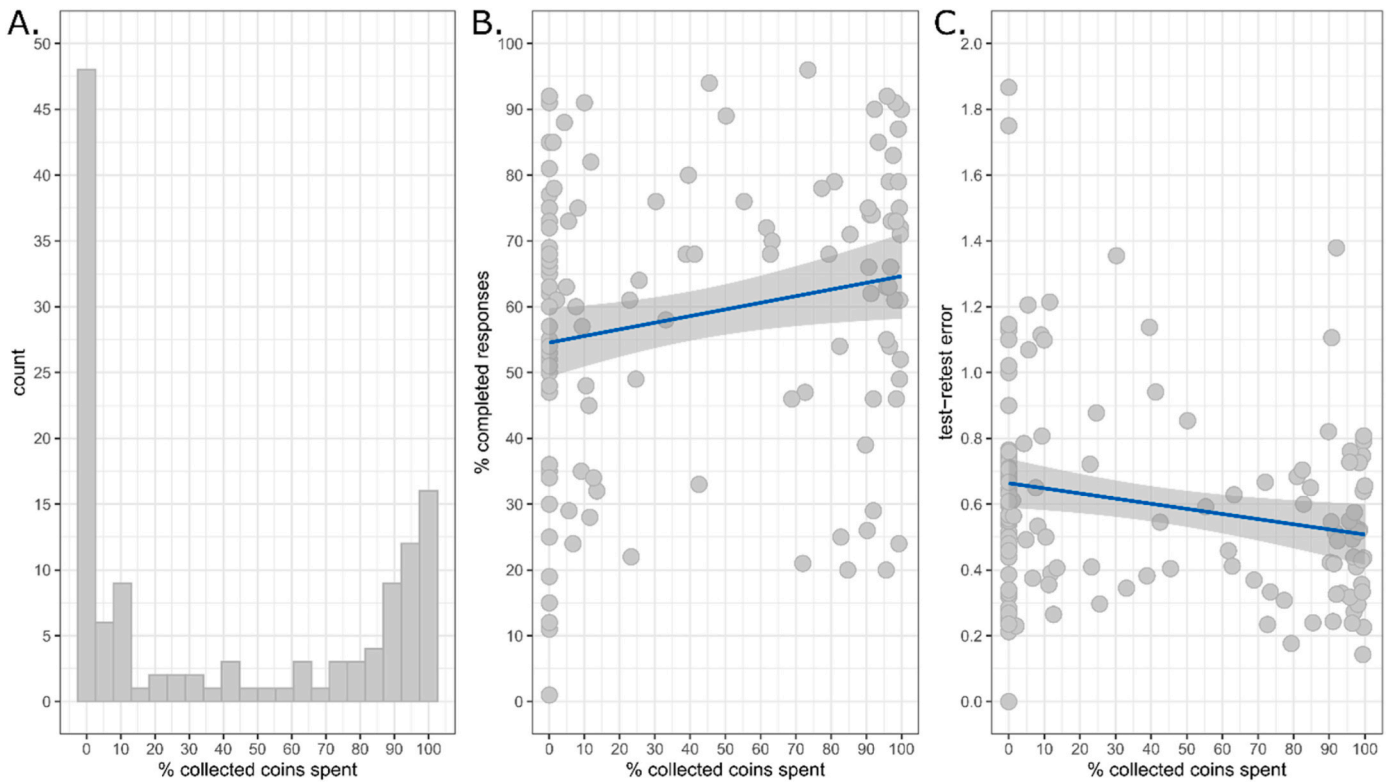

Fig. 3. Investigating the role of actively engaging with the gamified reward system. Panel A depicts the frequency distribution of participants that spent a particular percentage of collected coins on fun facts or personal graphs ($n = 128$). Panel B and C show how participants' response rate and reliabilities (somewhat) changed in function of their coin store purchases.

moderately rewarding ($M = 3.09$, $SD = 2.32$), a perception that was positively reflected in the number of coins they had actually spent ($r = .18$, $p = .038$).

7. Discussion

To maximize the information value of data collected in everyday life, a first prerequisite is that participants comply with the ESM protocol, meaning that they should frequently interact with the momentary surveys they receive. However, to do so on a regular basis is a burdensome endeavor (Eisele et al., 2020; Hasselhorn et al., 2022), and poor data quantity may undermine the ecological validity in ESM (Ram et al., 2017). In a real-life experiment, we aimed to prevent this potential predicament by instantaneously incentivizing people's response engagement by means of a gamification design. For every completed questionnaire, participants received a sum of virtual coins that functioned as a token to purchase in-app rewards like fun facts or personalized graphs.

7.1. Instantly gamifying survey completion in ESM and response engagement

Contrary to our initial expectations, simply adding this momentary gamification procedure to the ESM protocol was, on average, insufficient to promote higher response rates. Nevertheless, participants in the gamification condition who actively engaged with this reward strategy and exchanged a large proportion of their collected coins for virtual incentives showed higher response compliance. Instantly gamifying response behavior may therefore boost data quantity in ESM, but only to the extent participants show a natural motivation to engage with the study protocol. Upon further consideration, this finding somewhat aligns with the results reported by van Berkel, Goncalves and colleagues (2017), who found that higher response rates in their gamification condition were mainly the result of the participants' themselves initiating ESM surveys to improve their score on the leaderboard. In an ESM protocol that is fully time-contingent, motivated self-logging is evidently not possible, preventing significant differences in response compliance between control and gamification condition to appear. In the current study, participant motivation may have therefore manifested differently, namely the extent with which they fully interacted with the gamification procedure. Taken together, it is possible that instantly gamifying response behavior had a positive effect on the response rates of a subset of participants in the experimental condition. However, the direction of this effect remains speculative, as these analyses are refrained to a correlational level. It should also be noted that this effect was not very robust to outliers (see Supplemental Materials 4).

Second, participants' readiness to engage with the momentary notifications (or not) was also reflected in shorter response latencies. In an exploratory analysis, we found that people who actively purchased virtual incentives tended to respond to the time-based assessments faster. Consequently, they provided ratings closer to the randomly predefined sampling moments, which resulted in ESM assessments that were less biased by self-selection (Christensen et al., 2003). Because their data better approximated a truly random sample of moments in their everyday life, the ecological validity of their time series may have been higher (assuming of course that they did not change their daily routine to avoid missing assessments; Delespaul, 1995; Eisele et al., 2021).

7.2. Instantly gamifying survey completion in ESM and response quality

To obtain ESM time series with satisfactory information value, pursuing decent data quantity alone is not sufficient. That is, if participants interact with the momentary surveys on a regular basis, but do so in an unreliable way, the information value of collected data may remain net unchanged or be even worse. Here, we observed how a momentary

gamification design, on average, resulted in slightly more unreliable momentary ratings (as determined by larger test-retest discrepancies; Dejonckheere et al., 2022), which suggests that instantaneously rewarding response behavior generally resulted in somewhat lower-quality ESM data. However, this negative effect of gamification decreased when people more actively interacted with the coin store, again illustrating that participants' general motivation to engage with the study protocol was an important variable that moderated the effect of gamification on our response outcomes. Nevertheless, even individuals who proportionally spent an extensive amount of their collected coins still had slightly more unreliable momentary ratings than participants in the control condition who did not have access to a gamification procedure. These results contradict the findings of van Berkel et al. (2017), who observed how a gamified version of their ESM protocol produced momentary responses that were deemed of higher quality, as established via the evaluations of other participants in that study. A possible explanation for these diverging findings is that a gamified reward strategy may boost the quality of assessments that have a verifiable ground truth (i.e., higher inter-rater reliabilities for one-word location descriptions), but (somewhat) negatively impacts the quality of ratings of unverifiable subjective experiences (i.e., lower test-retest reliabilities for emotion ratings).

Remarkably, momentary test-retest reliabilities actually improved towards the end of the study, a temporal pattern that cannot be explained by the idea that participants found the surveys more burdensome as they were enrolled longer in the study. Instead, this finding was likely driven by the fact that ESM participants generally show decreased momentary response variance as the protocol continues (e.g., Eisele et al., 2023), which also reduces test-retest discrepancies. This underscores that the (test-retest) reliability scores of momentary measures may be an imperfect indicator of careless responding as a result of burden, since other processes or factors could also shape its presentation (e.g., changing contextual input, measurement reactivity, etc.; Dejonckheere et al., 2022).

For this reason, we also exploratory considered completion times of people's momentary measurements as a way to operationalize data quality. Submitting ESM responses very fast could be an indication of careless responding (Jaso et al., 2021; Schneider et al., 2018), and similar to previous studies (e.g., Eisele et al., 2023), we observed how participants submitted their responses faster towards the end of each day and towards the end of the study. Nevertheless, in the current study there was no evidence that our gamified ESM protocol stimulated unusually fast completion times. Following this quality indicator, there were no meaningful differences between standard and gamified ESM.

7.3. Persuasive incentive elements

Overall, we can conclude that participants found our gamified momentary reward scheme only moderately motivating. In behavioral terms, suboptimal engagement translated into a sizeable proportion of participants not purchasing anything from our coin store, even though survey completion provided them with virtual coins. Similarly, the coin bonus that increased with each subsequent completed assessment did not result in longer uninterrupted response streaks. In self-report terms, our manipulation check revealed that participants perceived our in-app incentives as only mildly rewarding. There were no real meaningful differences between extrinsically motivating fun facts and intrinsically motivating data visualizations. Regarding the latter, it is important to note that our graphs merely returned people's recorded responses without any substantive feedback. When these visual summaries are, for instance, supplemented with textual interpretations (e.g., *You seem to be particularly stressed when you are not doing well with respect to your "family" domain.*), it is possible that they do appeal to people's motivation to improve their self-insight (e.g., Hsieh et al., 2008).

Although our observed persuasive effects correspond with other effect sizes reported in the gamification literature (e.g., Mazeas et al., 2022), the question therefore arises whether including virtual game elements alone suffices to boost response compliance. When budget is short or ESM is applied in clinical settings, gamification could certainly be a viable incentive alternative if implemented correctly, but in standard scientific research, it may be better to complement this strategy with appropriate real-life incentive structures. Following this rationale, it may be worthwhile to investigate the impact of a gamification design where momentary in-app reinforcers directly correspond to real-world rewards (e.g., points earned with each completed assessment that can be exchanged for real gifts or money). In this way, the momentary mode of delivering incentives can effectively address survey completion at the within-person level (which explains most of the variance; Rintala et al., 2020; van Berkel et al., 2020), while optimally benefiting from the motivating qualities of real, more impactful rewards.

In sum, it is possible that the effect of our manipulation would have been more pronounced when our game-based incentive structure was more persuasive. This could also (partly) explain why our observed response rates were lower than those of other ESM studies.

7.4. Recommendations for incentivizing participants in future ESM research

Although the present results are not entirely equivocal when it comes to the effects of instantly gamifying response behavior in ESM on data quantity and quality, future ESM research could benefit from thinking more carefully about the impact of adopted incentive schemes. First, ESM researchers should consider whether participation compensation is *fixed* versus *incremental*. Although meta-analyses suggest that compliance-contingent reward schemes do not necessarily have higher response rates (Jones et al., 2019; Ottenstein & Werner, 2022), flat fees run the risk that some subjects simply respond to one momentary assessment, but still get reimbursed the full agreed renumeration (as was also observed a few times in the current study).

Next, in the case of incremental schemes, a second important decision is whether to apply *immediate* (token-based) versus *delayed* compensation strategies. According to traditional reinforcement theories (Miltenberger, 2015), immediate rewards have the advantage that the desired behavior easily stands out (and is thus likely to be repeated) compared to (one-time) full compensation after study completion. Any form of feedback during the ESM study (e.g., progress bars, compliance percentages) could stimulate participants to comply with the sampling protocol to the best of their possibilities. However, a potential downside is that this process information may move people's attention away from the actual study, mostly focusing on the encouragement mechanics that are used (Keusch & Zhang, 2017).

Third, it is imperative to take explicit notice of the incentivized target behavior. Current ESM studies predominantly aim to maximize data *quantity* with their reward strategies, while data *quality* is often ignored (perhaps because this is harder to establish; Jaso et al., 2021). However, simply because we stimulate participants to complete momentary surveys on a regular basis does not imply that they will do so in a reliable way. In fact, incentivizing one behavior may cause different (positive or negative) effects in other outcomes (Queirós & Pinto, 2022), and in the current study this was evident from the fact that exclusively rewarding response compliance paradoxically undermined the reliability of people's responses (i.e., larger test-retest discrepancies). To counter these detrimental effects, ESM researchers should therefore additionally take into account the quality of people's responses as a criterium in their reward schemes (e.g., by rewarding test-retest consistency with or without participants' knowledge). In that way, they can truly maximize the information value of the ESM time series they collect.

To further stimulate research into the potential of momentary reward schemes and/or gamification in ESM, we made various incentive/feedback features publicly available in m-Path's free version (Mestdagh et al., 2022). Researchers can fully customize a reward strategy that meets their specific study demands by tailoring various

building blocks such as progress bars, momentary awards, personal graphs and a coin store. Full set-up documentation can be found in our online manual (www.m-path.io/manual/).

7.5. Limitations and constraints on generality

The present findings should be interpreted in the light of some limitations. First, our experiment only evaluated the efficacy of one particular gamification procedure. Gamification is not a uniform concept, as researchers can install a motivating environment by combining game design elements in numerous ways, with different modes of delivery and with different content applications (Sailer et al., 2017). Consequently, conclusions about the general utility of gamification as a viable incentive strategy in ESM are unwarranted. On the one hand, the current study manipulated various game elements simultaneously (e.g., coin store, reward cabinet, etc.), making it difficult to isolate their individual contribution (although we made attempts to explore the impact of completion streaks and differentiated between the effects of intrinsic and extrinsic reward purchases). On the other hand, some other interesting game-based building blocks were not the focus of our investigation (e.g., compliance leaderboard, group challenges, etc.), but could be more or less effective in boosting response behavior in ESM. In addition, various scholars have advocated to reconsider the use of gamification as a one-fits-all incentive strategy (e.g., Rodrigues et al., 2021), and rather see value in optimally tailoring motivating game design elements for a particular individual (e.g., the HEXAD player type framework; Tondello et al., 2016). Evidence for this perspective emerges from the observation that our gamification application was only effective for a subgroup of participants in our sample.

Second, we did not explicitly evaluate participants' perceived burden with the ESM protocol (e.g., Eisele et al., 2020; Hasselhorn et al., 2022). This information would have provided us with a better understanding of the active mediating mechanisms through which momentary incentive strategies impact response outcomes. That is, it is equally possible that the added gamification elements resulted in an improved user experience of the mobile ESM app, or that they incited other motives to comply. In general, the field could benefit from focusing more on *how* gamification is effective, as opposed to *whether* (Sailer & Homner, 2020).

Finally, the provision of a real-life incentive (i.e., money or course credits) may have mitigated some of the current study effects. Although a real-world compensation was essential to make our study as appealing to take part in as other concurrent studies, not including this perhaps would have left the differential effects between gamification and control condition to be more outspoken.

8. Conclusion

Real-time incentivizing the completion of momentary surveys in ESM by means of a gamification procedure could lead to higher response rates for some participants. However, merely rewarding the act of completing an assessment, and therefore exclusively trying to maximize data quantity, may threaten the ecological validity of ESM research, as this strategy paradoxically pushes participants to provide less reliable responses (undermining data quality). We therefore advise researchers to include both measures of response quantity and quality in the reward schemes of future ESM studies.

CRediT authorship contribution statement

Egon Dejonckheere: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Stijn Verdonck:** Writing – review & editing, Supervision, Software, Resources, Methodology, Investigation, Conceptualization. **Joren Andries:** Writing – review & editing, Investigation, Data curation. **Natalie Röhrig:** Writing – review & editing, Investigation, Data curation. **Maarten Piot:** Writing – review & editing, Software. **Ghijs Kilani:** Writing – review & editing, Software. **Merijn Mestdagh:** Writing – review & editing, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: ED, SV and MM co-founded m-Path, the commercial experience sampling software that was used in this study. MP and GK are employees in the same company. The other authors declare no conflict of interest.

Data availability

All data and materials can be accessed at the Opens Science Framework: https://osf.io/ctznf/.

Acknowledgments

The current study was supported by the Research Fund of KU Leuven (C14/19/054). Merijn Mestdagh and Egon Dejonckheere are postdoctoral research fellows supported by the Fund for Scientific Research, Flanders (FWO; 1256221N, 1210621N). The funding sources had no involvement in the collection of the data, analyses or results reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2024.108360.

References

Bai, S., Foon Hew, K., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, Article 100322.

Bittner, J. V., & Shipper, J. (2014). Motivational effects and age differences of gamification in product advertising. *Journal of Consumer Marketing, 31*, 391–400.

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments, 24*, 1162–1175.

Carlier, C., Niemeijer, K., Mestdagh, M., Bauwens, M., Vanbrabant, P., Geurts, L., et al. (2022). In search of state and trait emotion markers in mobile-sensed language: Field study. *JMIR Mental Health, 9*, Article e31724.

Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Kaschub, C. (2003). A practical guide to experience-sampling procedures. *Journal of Happiness Studies, 4*, 53–78.

Conner, T. S., & Lehman, B. J. (2012). Getting started: Launching a study in daily life. In *Handbook of research methods for studying daily life* (pp. 89–107). The Guilford Press.

Courvoisier, D. S., Eid, M., & Lischetzke, T. (2012). Compliance to a cell phone-based ecological momentary assessment study: The effect of time and personality characteristics. *Psychological Assessment, 24*, 713–720.

De Croon, R., Geuens, J., Verbert, K., & Vanden Abeele, V. (2021). A systematic review of the effect of gamification on adherence across disciplines. In *HCI in Games: Experience design and game mechanics* (pp. 168–184). Springer International Publishing.

Dejonckheere, E., Demeyer, F., Geusens, B., Piot, M., Tuerlinckx, F., Verdonck, S., et al. (2022). Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological Assessment, 34*, 1138–1154.

Dejonckheere, E., & Erbas, Y. (2021). Designing an experience sampling study. In *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (pp. 33–70). Center for Research on Experience sampling and Ambulatory methods Leuven.

Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., et al. (2018). The bipolarity of affect and depressive symptoms. *Journal of Personality and Social Psychology, 114*, 323–341.

Delespaul, P. A. E. G. (1995). *Assessing schizophrenia in daily life: The experience sampling method*. Universitaire pers Maastricht.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. Using game-design elements in non-gaming contexts. *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 2425–2428.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., et al. (2020). The effects of sampling frequency and questionnaire length on perceived burden,

compliance, and careless responding in experience sampling data in a student population. *Assessment, 29*, 136–151.
Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., et al. (2023). A mixed-method investigation into measurement reactivity to the experience sampling method: The role of sampling protocol and individual characteristics. *Psychological Assessment, 35*, 68–81.
Eisele, G., Vachon, H., Myin-Germeys, I., & Viechtbauer, W. (2021). Reported affect changes as a function of response delay: Findings from a pooled dataset of nine experience sampling studies. *Frontiers in Psychology, 12*, Article 580684.
Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., et al. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry, 26*, 105–109.
Fuller-Tyszkiewicz, M., Skouteris, H., Richardson, B., Blore, J., Holmes, M., & Mills, J. (2013). Does the burden of the experience sampling method undermine data quality in state body image research? *Body Image, 10*, 607–613.
Harari, G. M., Müller, S. R., Mishra, V., Wang, R., Campbell, A. T., Rentfrow, P. J., et al. (2017). An evaluation of students' interest in and compliance with self-tracking methods: Recommendations for incentives based on three smartphone sensing studies. *Social Psychological and Personality Science, 8*, 479–492.
Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice, 10*, 169–207.
Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2022). The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment. *Behavior Research Methods, 54*, 1541–1558.
Hektner, J. M., & Csikszentmihalyi, M. (2002). The experience sampling method: Measuring the context and the content of lives. In *Handbook of environmental psychology* (pp. 233–243). John Wiley & Sons, Inc.
Hsieh, G., Li, I., Dey, A., Forlizzi, J., & Hudson, S. E. (2008). Using visualizations to increase compliance in experience sampling. In *Proceedings of the 10th international Conference on Ubiquitous computing* (pp. 164–167).
Hursh, S. R. (1984). Behavioral economics. *Journal of the Experimental Analysis of Behavior, 42*, 435–452.
Jaso, B. A., Kraus, N. I., & Heller, A. S. (2021). Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring. *Psychological Methods, 27*, 958–981.
Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., et al. (2019). Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction, 114*, 609–619.
Kalokerinos, E. K., Moeck, E. K., Rummens, K., Meers, K., & Mestdagh, M. (2022). Ready for the worst? Negative affect in anticipation of a stressor does not protect against affective reactivity. *Journal of Personality, 91*, 1123–1139.
Kazdin, A. (1977). *The token economy: A review and evaluation*. Springer US.
Keusch, F., & Zhang, C. (2017). A review of issues in gamified surveys. *Social Science Computer Review, 35*, 147–166.
Long, B., Simson, J., Buxó-Lugo, A., Watson, D. G., & Mehr, S. A. (2023). How games can make behavioural science better. *Nature, 613*, 433–436.
Ludwigs, K., Lucas, R., Veenhoven, R., Richter, D., & Arends, L. (2020). Can happiness apps generate nationally representative datasets? A case study collecting data on people's happiness using the German socio-economic panel. *Applied Research in Quality of Life, 15*, 1135–1149.
Lynn, P. (2001). The impact of incentives on response rates to personal interview surveys: Role and perceptions of interviewers. *International Journal of Public Opinion Research, 13*, 326–336.
Mölsä, M. E., Lax, M., Korhonen, J., Gumpel, T. P., & Söderberg, P. (2022). The experience sampling method in monitoring social interactions among children and adolescents in school: A systematic literature review. *Frontiers in Psychology, 13*, Article 844698.
Mazeas, A., Duclos, M., Pereira, B., & Chalabaev, A. (2022). Evaluating the effectiveness of gamification on physical activity: Systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research, 24*, Article e26779.
Messiah, A., Grondin, O., & Encrenaz, G. (2011). Factors associated with missing data in an experience sampling investigation of substance use determinants. *Drug and Alcohol Dependence, 114*, 153–158.
Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology research: Current achievements and future ambitions. *Current Opinion in Psychology, 41*, 1–8.
Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Tuerlinckx, F., Kuppens, P., et al. (2022). m-Path: An easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. *Frontiers in Digital Metal Health, 5*, Article 182175.
Miltenberger, R. G. (2015). *Behavior modification: Principles and procedures* (6th ed.). Cengage Learning.
Morren, M., van Dulmen, S., Ouwerkerk, J., & Bensing, J. (2009). Compliance with momentary pain measurement using electronic diaries: A systematic review. *European Journal of Pain, 13*, 354–365.
Musthag, M., Raij, A., Ganesan, D., Kumar, S., & Shiffman, S. (2011). Exploring micro-incentive strategies for participant compensation in high-burden studies. In *Proceedings of the 13th International Conference on Ubiquitous computing* (pp. 435–444).
Newcomb, M. E., Swann, G., Estabrook, R., Corden, M., Begale, M., Ashbeck, A., et al. (2018). Patterns and predictors of compliance in a prospective diary study of substance use and sexual behavior in a sample of young men who have sex with men. *Assessment, 25*, 403–414.
Ono, M., Schneider, S., Junghaenel, D. U., & Stone, A. A. (2019). What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. *Journal of Medical Internet Research, 21*, Article e11398.
Ottenstein, C., & Werner, L. (2022). Compliance in ambulatory assessment studies: Investigating study and sample characteristics as predictors. *Assessment, 29*, 1765–1776.
Oyibo, K., Orji, R., & Vassileva, J. (2017). The influence of culture in the effect of age and gender on social influence in persuasive technology. In *Adjunct Publication of the 25th Conference on user modeling, Adaptation and Personalization* (pp. 47–52).
Piasecki, T. M., Hufford, M. R., Solhan, M., & Trull, T. J. (2007). Assessing clients in their natural environments with electronic diaries: Rationale, benefits, limitations, and barriers. *Psychological Assessment, 19*, 25–43.
Queirós, R., & Pinto, M. (2022). The dangers of gamification. In *Advanced research in Technologies, information, Innovation and Sustainability* (pp. 151–162). Nature Switzerland: Springer.
Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: Considerations for design and analysis. *Research in Human Development, 14*, 253–270.
Rescorla, R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning: Current research and theory*. NY: Appleton Century Crofts.
Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment, 31*, 226–235.
Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. *Psychiatry Research, 286*, Article 112896.
Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons, 58*, 411–420.
Rodrigues, L., Toda, A. M., Palomino, P. T., Oliveira, W., & Isotani, S. (2021). Personalized gamification: A literature review of outcomes, experiments, and approaches. In *8th International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 699–706).
Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380.
Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*, 77–112.
Schneider, S., May, M., & Stone, A. A. (2018). Careless responding in internet-based quality of life assessments. *Quality of Life Research, 27*, 1077–1088.
Scollon, N. C., Prieto, C.-K., & Diener, E. (2009). Experience sampling: Promises and pitfalls, strength and weaknesses. In *Assessing well-being: The collected works of Ed Diener* (pp. 157–180). Netherlands: Springer.
Silvia, P. J., Kwapil, T. R., Eddington, K. M., & Brown, L. H. (2013). Missed beeps and missing data: Dispositional and situational predictors of nonresponse in experience sampling research. *Social Science Computer Review, 31*, 471–481.
Sokolovsky, A. W., Mermelstein, R. J., & Hedeker, D. (2014). Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine & Tobacco Research, 16*, 351–358.
Soyster, P. D., Bosley, H. G., Reeves, J. W., Altman, A. D., & Fisher, A. J. (2019). Evidence for the feasibility of person-specific ecological momentary assessment across diverse populations and study designs. *Journal for Person-Oriented Research, 5*, 53–64.
Stone, A., Broderick, J., Schwartz, J., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain, 104*, 343–351.
Stone, A. A., Kessler, R. C., & Haythornthwaite, J. A. (1991). Measuring daily events and experiences: Decisions for the researcher. *Journal of Personality, 59*, 575–607.
Sun, J., Rhemtulla, M., & Vazire, S. (2021). Eavesdropping on missing data: What are university students doing when they miss experience sampling reports? *Personality and Social Psychology Bulletin, 47*, 1535–1549.
Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The gamification user types hexad scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human interaction in play* (pp. 229–243).
Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research, 21*, Article e14475.
van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys, 50*, 93.
van Berkel, N., Goncalves, J., Hosio, S., & Kostakos, V. (2017). Gamification of mobile experience sampling improves data quality and quantity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*, 107.
van Berkel, N., Goncalves, J., Hosio, S., Sarsenbayeva, Z., Velloso, E., & Kostakos, V. (2020). Overcoming compliance bias in self-report studies: A cross-study analysis. *International Journal of Human-Computer Studies, 134*, 1–12.
van Berkel, N., Goncalves, J., Koval, P., Hosio, S., Dingler, T., Ferreira, D., et al. (2019). Context-informed scheduling and analysis: Improving accuracy of mobile self-reports. In *Proceedings of the 2019 CHI Conference on Human factors in computing systems* (pp. 1–12).
van Berkel, N., & Kostakos, V. (2021). Recommendations for conducting longitudinal experience sampling studies. In *Advances in Longitudinal HCI research* (pp. 59–78). Springer International Publishing.
van Roekel, E., Keijsers, L., & Chung, J. M. (2019). A review of current ambulatory assessment studies in adolescent samples and practical recommendations. *Journal of Research on Adolescence, 29*, 560–577.
von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM, 51*, 58–67.

von Klipstein, L., Riese, H., van der Veen, D. C., Servaas, M. N., & Schoevers, R. A. (2020). Using person-specific networks in psychotherapy: Challenges, limitations, and how we could use them anyway. *BMC Medicine, 18*, 345.

Wen, C. K. F., Schneider, S., Stone, A. A., & Spruijt-Metz, D. (2017). Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. *Journal of Medical Internet Research, 19*, e132.

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.

Wrzus, C., & Neubauer, A. B. (2022). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment, 30*, 825–846.