The observer in the Pocket: Investigating Moderators and Impact of Measurement Reactivity in Digital Dietary Intake Monitoring

Dissertation

For the award of the academic degree Dr. phil. (Doctor of Philosophy)
at the Faculty of Life Sciences:
Food, Nutrition and Health
at the University of Bayreuth
Germany

Submitted by:
Anila Allmeta
Kavaje, Albania

04.05.2026

D I S S E R T A T I O N
zur Erlangung des akademischen Grades

Doctor philosophiae
(Dr. phil.)

eingereicht an der
Fakultät für Lebenswissenschaften: Lebensmittel, Ernährung und Gesundheit der Universität Bayreuth

von Anila Allmeta

Die vorliegende Arbeit wurde in der Zeit von November 2020 bis Mai 2026 am Lehrstuhl für Public Health Nutrition der Universität Bayreuth unter der Betreuung von Jun.-Prof. Dr. Laura König begonnen und nach deren Wechsel an die Universität Wien unter ihrer Betreuung abgeschlossen.

Art der Dissertation:	kumulativ
Dissertation eingereicht am:	04. Mai 2026
Zulassung durch Promotionskommission:	20. Mai 2026
Wissenschaftliches Kolloquium:	17. Juli 2026

Präsident der Universität Bayreuth

Prof. Dr. Stefan Leible

Dekanin der Fakultät für Lebenswissenschaften: Lebensmittel, Ernährung und Gesundheit der Universität Bayreuth

Prof. Dr. Corina Vlot-Schuster

Gutachter/innen

Prof. Dr. Laura König
Prof. Dr. Peter von Philipsborn
Prof. Dr. Janin Henkel-Oberländer (Vorsitz)
Jun.-Prof. Dr. Tim Dorlach

D I S S E R T A T I O N
for the academic degree of

Doctor philosophiae
(Dr. phil.)

submitted to the
Faculty of Life Sciences: Food, Nutrition, and Health
at the University of Bayreuth

by Anila Allmeta

This dissertation was carried out between November 2020 and May 2026. It was initiated at the Chair of Public Health Nutrition at University of Bayreuth under the supervision of Jun.-Prof. Dr. Laura König and completed under her supervision following her appointment to University of Vienna.

Type of dissertation:	cumulative
Dissertation submitted on:	04 May, 2026
Approval by the doctoral committee:	20 May, 2026
Scientific colloquium:	17 July, 2026

President of the University of Bayreuth

Prof. Dr. Stefan Leible

Dean of the Faculty of Life Sciences: Food, Nutrition, and Health
at the University of Bayreuth

Prof. Dr. Corina Vlot-Schuster

Reviewers

Prof. Dr. Laura König
Prof. Dr. Peter von Philipsborn
Prof. Dr. Janin Henkel-Oberländer (Chair)
Jun.-Prof. Dr. Tim Dorlach

Acknowledgments

Firstly, I would like to express my deepest gratitude to my brilliant supervisor Prof. Dr. Laura König. Your guidance was the cornerstone of my growth; I am very thankful for the intellectual empowerment and for being such a reliable guide who celebrated my success as your own. Thank you for providing me with the opportunities to grow as a researcher and connect with numerous inspiring individuals. I will always cherish our time together!

I also sincerely thank my second supervisor and committee chair, Prof. Dr. Janin Henkel-Oberländer, for stepping in at a critical moment and overseeing the final stages with such care and professionalism.

This work was a collaborative milestone, and I am indebted to my co-authors, Prof. Dr. Stephen Sutton, Prof. Dr. Danielle Arigo, Dr. Andreas Neubauer, and Dr. Alea Ruf, for the stimulating discussions that brought our projects to fruition. To all my colleagues at the Public Health Nutrition Chair: thank you for the vibrant and intellectually stimulating environment. The PHN team holds a special place in my heart. Additionally, a huge thank you to Mirna for the long talks over delicious Kuchen in Kulmbach; those moments made this journey very special.

On a personal note, a special thanks to my family: my brother Altin, my endless source of love and support; my mother, Zegjine, my inspiration and everyday S-HERO; and my father, Shkëlqim, my angel up in heaven. Thank you for the endless encouragement and all the sacrifices you made to support my education. Finally, words cannot fully express my gratitude to my anchor and biggest supporter, my husband Suad. Your patience, belief in me, and infinite love made this journey meaningful! I could not have done this without you by my side so thank you for taking care of US. We made it!

Ju dua pafundësisht shumë!

Table of Contents

Preliminary publications of this dissertation

A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour

König, L. M., **Allmeta, A.,** Christlein, N., Van Emmenis, M., & Sutton, S. (2022). A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour. *Health Psychology Review*, *16*(4), 551–575. https://doi.org/10.1080/17437199.2022.2047096

Contribution: SS and LK conceived of the study and developed the research question. LK developed the search strategy with input from SS, **AA** and MVE. Searches were conducted by LK; screening was conducted by LK, **AA**, MVE and NC; data was extracted by LK and NC; study quality was appraised by **AA**, LK and MVE. Data was analyzed and interpreted, and the manuscript was written by LK with input from all authors.

The same, only different: Smartphone-based dietary Ecological Momentary Assessment tools vary in complexity, usability, and active information processing.

Allmeta, A., Sutton, S., & König, L. M. (2026). The same, only different: Smartphone-based dietary Ecological Momentary Assessment tools vary in complexity, usability and active information processing. *British Journal of Health Psychology*, *31*(1), e70057. https://doi.org/10.1111/bjhp.70057

Contribution: **AA**: conceptualization, data curation, writing – original draft, formal analysis, investigation, methodology, project administration; SS: conceptualization, methodology, writing-review & editing; LK: conceptualization, methodology, writing-review & editing.

Easy or difficult? Investigating perceived ease of changing eating and physical activity behaviors

Allmeta, A., Arigo, D., & König, M. L. (2026). Easy or difficult? Investigating perceived ease of changing eating and physical activity behaviors. *Applied Psychology Health and Well-Being*, *18*(1), e70124. https://doi.org/10.1111/aphw.70124

Contribution: **AA**: conceptualization, methodology, formal analysis, investigation, data curation, writing-original draft; DA: conceptualization, writing-review & editing; LK: conceptualization, methodology, writing-review & editing, supervision.

Measurement reactivity in smartphone -based dietary ecological momentary assessment

Allmeta, A., Ruf, A., Neubauer, A., & König, L. M. (2026). Measurement reactivity in smartphone-based dietary ecological momentary assessment. Retrieved from osf.io/preprints/psyarxiv/uw2gk_v1

Contribution: **AA**: conceptualization, methodology, formal analysis, investigation, data curation, writing-original draft; AR: software, investigation, writing-review & editing; AN: conceptualization, writing-review & editing; LK: conceptualization, methodology, writing-review & editing, supervision.

Conference Presentations

Allmeta, A., Ruf, A., Neubauer, A., König, L. (2023). Measurement Reactivity in Dietary Ecological Momentary Assessment: Does Familiarity with Tracking Make a Difference? Presentation in the 37th Conference of the European Health Psychology, Bremen, Germany

Allmeta, A., Sutton, S., König, L. (2022). Do more complex food tracking apps induce greater changes in eating-related cognitions, intentions, and behaviour? Presentation in the Deutschen Gesellschaft für Verhaltensmedizin (DGVM) Conference, Salzburg, Austria

Allmeta, A., Arigo, D., König, L. (2022). Easy or difficult? Investigating perceived ease to change aspects of eating behaviour and physical activity. Presentation in the 36th Conference of the European Health Psychology, Bremen, Germany

List of Abbreviations

EMA	Ecological Momentary Assessment
AA	Ambulatory Assessment
MR	Measurement Reactivity
AIP	Active Information Processing
SRM	Self-Regulation Model
PA	Physical Activity
SB	Sedentary Behavior
EI	Energy Intake
FV	Fruits and Vegetables
OSF	Open Science Framework

Abstract (German)

Hintergrund: Die grundlegende Prämisse der verhaltenswissenschaftliche Forschung beruht auf der ökologischen Validität der erfassten Daten und geht davon aus, dass das in einer Studie gemessene Verhalten dem Verhalten außerhalb des Studienkontexts entspricht. Mit dem Wandel des Forschungsfeldes hin zu intensiven Längsschnittstudien wie der ökologischen Momentaufnahme (Ecological Momentary Assessment; EMA) und digitalen Bewertungen entsteht ein „Validitätsparadoxon". Digitale Tools ermöglichen zwar präzisere Messungen als herkömmliche Fragebögen, doch die erhöhte Transparenz und die ständigen Feedbackschleifen können unbeabsichtigt die Reaktivitätseffekte der Messungen verstärken. Dieses Phänomen, definiert als die unbeabsichtigte Verhaltens- oder kognitive Veränderung, die sich aus dem Beurteilungsprozess selbst ergibt, stellt eine Bedrohung für die Integrität von „naturalistischen" Daten dar. Während die Reaktivität bei körperlicher Aktivität bereits erforscht wurde, ist ihr Einfluss auf die digitale Ernährungsbewertung noch weitgehend unerforscht. Das übergeordnete Ziel dieser Dissertation war die detaillierte Untersuchung der Messreaktivität, ihrer zugrunde liegenden psychologischen Mechanismen und der potenziellen Einflussfaktoren, die ihr Ausmaß im Kontext der smartphonebasierten Ernährungsüberwachung bestimmen. Zu diesem Zweck schlagen wir ein Interaktionsmodell der Reaktivität vor, in dem die Verhaltensauswirkung aus einer Interaktion zwischen den Möglichkeiten des Werkzeugs (Belastung, Klarheit, Feedback), dem kognitiven Engagement des Benutzers (aktive Informationsverarbeitung; AIP) und der wahrgenommenen Fähigkeit (wahrgenommene Leichtigkeit der Veränderung) mit dem Zielverhalten entsteht.

Methodischer Ansatz: Um dieses Ziel zu erreichen, wurde ein systematischer vierstufiger Ansatz umgesetzt. Studie 1 legte die empirische Grundlage durch eine systematische

Übersichtsarbeit und Metaanalyse (k = 33) zur digitalen „im Moment" Reaktivität im gesamten Gesundheitsverhalten. Studie 2 verlagerte den Fokus auf das Bewertungsinstrument selbst und verwendete ein experimentelles Modell, um zu evaluieren, wie die architektonische Komplexität des Interface des EMA für Ernährung die aktive Informationsverarbeitung (AIP) und die wahrgenommene Benutzerfreundlichkeit beeinflusst. Studie 3 (n = 1.072) untersuchte die Verhaltenskomponente und erstellte eine Hierarchie der wahrgenommenen Veränderungsleichtigkeit über 13 Aspekte im Zusammenhang mit dem Essverhalten und 8 Aspekte der körperlichen Aktivität hinweg, um herauszufinden, welche Ziele am anfälligsten für Messungenechos sind. Schließlich wurde in Studie 4 ein achttägiges ereignisabhängiges EMA-Protokoll (n = 82) verwendet, um das vorgeschlagene Interaktionsmodell in einem realen Umfeld zu testen. Dabei wurde untersucht, wie Werkzeugkomplexität, Verhaltensplastizität und individuelle Merkmale zusammenwirken, um die Messreaktivität zu beeinflussen.

Ergebnisse: Die systematische Überprüfung (Studie 1) bestätigte, dass digitale Reaktivität ein vorhandenes, wenn auch subtiles Phänomen ist, das durch kleine signifikante Effekte gekennzeichnet ist (Cohen´s *d* Range: 0,06 bis 0,28). Entscheidend war die Erkenntnis, dass die Reaktivität nicht einheitlich ist; Verhaltensweisen wie körperliche Aktivität und Alkoholkonsum zeigten deutliche Veränderungen, während es für die Nahrungsaufnahme konkrete Belege fehlten. Darüber hinaus wurden mehrere potenzielle Moderatoren der Messreaktivität aufgezeigt, wie beispielsweise (1) AIP; (2) Leichtigkeit der Veränderung; (3) Absichten zur Verhaltensänderung; (4) Vertrautheit mit der Nachverfolgung. Studie 2 spiegelte die Wahrnehmungen der Teilnehmenden hinsichtlich verschiedener EMA-Tools zur Erfassung der Ernährung wider: Sie wurden als mehr oder weniger komplex wahrgenommen. Komplexere Tools, die ein hohes Maß an Interaktion und detaillierter Dokumentation erforderten, führten laut Einschätzung der Teilnehmenden zu einem signifikant höheren AIP-

Wert. Darüber hinaus wurde eine „U-förmige“ Beziehung festgestellt: Während ein Mindestmaß an AIP notwendig ist, um Selbstreflexion auszulösen, verringert übermäßige Komplexität die Benutzerfreundlichkeit und kann das Engagement der Teilnehmer gefährden. Studie 3 lieferte eine psychologische „Fahrplan“ für Reaktivität, indem sie aufzeigte, dass Individuen bestimmte Aspekte von Verhaltensweisen als leichter oder schwieriger zu ändern wahrnehmen. Additive Verhaltensweisen (z. B. Erhöhung des Obst- und Gemüsekonsums) wurden als leichter zu ändern wahrgenommen als subtraktive oder hemmende Verhaltensweisen (z. B. Reduzierung des Konsums ungesunder Snacks). Eine hohe wahrgenommene Leichtigkeit der Veränderung erwies sich als Prädiktor für die Verhaltensabsicht, was darauf hindeutet, dass eine Reaktion am ehesten dann auftritt, wenn die „Eintrittsbarriere“ für eine Veränderung niedrig ist. Studie 4 integrierte die Ergebnisse der Studien 1 bis 3 und bestätigte, dass die Ernährungs-EMA tatsächlich reaktiv ist, insbesondere für ansatzorientiertes gesundes Verhalten. Teilnehmer mit hoher Gesundheitskompetenz, die komplexe Tools zur Erfassung von Verhaltensweisen nutzten, die ihrer Meinung nach leicht zu ändern waren, zeigten die höchste Reaktivität, insbesondere beim Obst- und Gemüsekonsum. Im Gegensatz dazu blieben präventive oder hemmende Verhaltensweisen, wie der Konsum von Snacks, resistent gegen kurzfristige reaktive Effekte, was wahrscheinlich auf ihren tief verwurzelten oder automatischen Charakter zurückzuführen ist.

Implikationen: Diese Dissertation bringt das Fachgebiet voran, indem sie über die binäre Frage hinausgeht, ob Reaktivität existiert, und hin zu einem „Person-Verhalten-Tool-Interaktion“-Ansatz geht. Theoretisch bieten die Ergebnisse eine entscheidende Weiterentwicklung des Selbstregulierungsmodells (SRM). Während das SRM darauf hindeutet, dass jede Überwachung eine Diskrepanz-Reduktionsschleife auslöst, zeigt diese Arbeit, dass die Schleife selektiv ist und eine „fördernde“ Selbstregulierung (das Streben nach einem positiven Ziel) gegenüber einer „präventiven“ Regulierung in kurzfristigen digitalen

Ausbrüchen bevorzugt. Darüber hinaus bietet die Einführung von AIP als granularer Mechanismus eine neue Perspektive für das Verständnis, wie die kognitive Tiefe einer Bewertung die Verhaltensanpassung beeinflusst. Für die Forschungsgemeinschaft dienen diese Erkenntnisse als methodischer Auftrag für „reaktivitätsbewusste“ Designs. Forscher sollten „Burn-in“-Perioden einführen (sorgfältige Überwachung früher Datenpunkte) und den AIP routinemäßig als Standardprozessvariable messen, um den „Überwachungseffekt“ vom „Tooleffekt“ zu trennen. In der klinischen Praxis sollte die Messreaktivität nicht als rein psychometrischer Bias betrachtet werden, der beseitigt werden muss, sondern als „kostenloser“ digitaler Anstoß. Die von einer App erfassten optimierten Daten können die maximale Veränderungsfähigkeit eines Patienten unter Unterstützung darstellen und Praktiker eine Grundlage für die erreichbare Leistung bieten.

Zusammenfassend kommt diese Dissertation zu dem Schluss, dass menschliches Gesundheitsverhalten von Natur aus beobachtungssensitiv ist. Durch den Nachweis, dass Reaktivität ein Produkt der Werkzeugkomplexität, der Verhaltensplastizität sowie des individuellen kognitiven Engagements und spezifischer Merkmale ist, bietet diese Arbeit einen vielseitigen Rahmen für zukünftige ambulante Forschungen und stellt sicher, dass während wir weiterhin das Leben in Echtzeit messen, tun wir dies mit einem ausgefeilten Verständnis dafür, wie der Beobachter das Beobachtete unweigerlich verändert.

Abstract (English)

Background: The foundational premise of behavioral research relies on the ecological validity of captured data, assuming that the behavior measured in a study equals the behavior shown outside the study context. As the field shifts toward intensive longitudinal designs such as Ecological Momentary Assessment (EMA) and digital assessments, a “validity paradox” emerges. While digital tools provide more precise measurements than traditional questionnaires, the heightened transparency and constant feedback loops may inadvertently amplify measurement reactivity effects. This phenomenon, defined as the unintended behavioral or cognitive shift resulting from the assessment process itself, poses a threat to the integrity of “naturalistic” data. While reactivity has been documented in physical activity, its presence in digital dietary assessment remains under-investigated. The overarching aim of this dissertation was to provide a granular investigation into the presence of measurement reactivity, its underlying psychological mechanisms, and the potential moderators that dictate its magnitude within the context of smartphone-based dietary monitoring. As a result, we suggest an interaction model of reactivity, in which behavioral impact emerges from an interaction between the tool’s affordances (burden, clarity, feedback), the user’s cognitive engagement (active information processing; AIP) and perceived ability (perceived ease of change) with the targeted behavior.

Method: To address this aim, a four-stage systematic trajectory was implemented. Study 1 established the empirical foundation through a systematic review and meta-analysis (k = 33) of digital in-the-moment reactivity across health behaviors. Study 2 shifted the focus to the assessment tool itself, using an experimental mock-up design to evaluate how the architectural complexity of dietary EMA interface influences Active Information Processing (AIP) and perceived usability. Study 3 (n = 1,072) investigated the behavioral component, establishing a hierarchy of the perceived ease of change across 13 eating-related and 8 physical activity aspects to identify which targets are most susceptible to measurement reactivity. Finally, Study 4, utilized an eight-day event-contingent EMA protocol (n = 82) to test the proposed interaction model in a real-world setting, examining how tool complexity, behavioral plasticity, and individual traits converge to drive measurement reactivity.

Results: The systematic review (Study 1) confirmed that digital reactivity is a present, albeit subtle, phenomenon, characterized by small significant effects (Cohen´s d range: 0.06 to 0.28). Crucially, it revealed that reactivity is not uniform; behaviors like physical activity and alcohol consumption showed pronounced shifts, whereas dietary intake lacked dedicated evidence. Furthermore, it outlined several potential moderators of measurement reactivity such as (1) AIP; (2) ease of change; (3) intentions to change behavior; (4) familiarity with tracking. Study 2 reflected participants´ perceptions on different dietary EMA tools: perceived as more or less complex. More complex tools which required high levels of interaction and detailed tracking were also perceived to induce significantly higher AIP. Additionally, a "U-shaped" relationship was identified: while a minimum level of AIP is necessary to trigger self-reflection, excessive complexity reduced usability and could threaten participant engagement.

Study 3 provided a psychological "roadmap" for reactivity by showing that individuals perceive certain aspects of behaviors as easier or more difficult to change. Additive behaviors (e.g., increasing fruit/vegetable intake) were perceived as easier to change than subtractive or inhibitory behaviors (e.g., reducing unhealthy snack consumption). High perceived ease of change emerged as a predictor of behavioral intention, suggesting that reactivity is most likely to occur when the "entry barrier" to change is low. Study 4 integrated findings from studies 1 to 3, confirming that dietary EMA is indeed reactive, particularly for approach-oriented healthy behaviors. Participants with previous experience in using complex tools to record

behaviors they perceived as easy to change showed the highest levels of reactivity, specifically in fruit and vegetable consumption. In contrast, preventive or inhibitory behaviors, such as snack consumption, remained resistant to short-term reactive effects, likely due to their entrenched or automatic nature.

Implications: This dissertation advances the field by moving beyond the binary question of whether reactivity exists toward a "Person-Behavior-Tool Interaction" approach. Theoretically, the findings offer a critical refinement of the Self-Regulation Model (SRM). While the SRM suggests that any monitoring triggers a discrepancy-reduction loop, this work indicates that the loop is selective, favoring "promotive" self-regulation (striving for a positive goal) over "preventive" regulation in short-term digital bursts. Furthermore, the introduction of AIP as a granular mechanism provides a new lens for understanding how the cognitive depth of an assessment drives behavioral adjustment.

For the research community, these findings serve as a methodological mandate for "reactivity-aware" designs. Researchers should implement "burn-in" periods (cautiously monitoring early data points) and routinely measure AIP as a standard process variable to disentangle the "monitoring effect" from the "tool effect". In clinical practice, measurement reactivity should not be viewed as a purely psychometric bias to be eliminated, but as a "free" digital nudge. The optimized data captured by an app may represent a patient's maximal capacity for change under support, offering practitioners a baseline for achievable performance.

In summary, this dissertation concludes that human health behavior is inherently observation-sensitive. By demonstrating that reactivity is a product of tool complexity, behavioral plasticity, and individual cognitive engagement and specific characteristics, this work provides a versatile framework for future ambulatory research, ensuring that as we continue to measure life in real-time, we do so with a sophisticated understanding of how the observer inevitably transforms the observed.

1. General Introduction

1.1. Description

A fundamental, yet often scrutinized, assumption in health psychology is that behavior observed within the structured confines of a research protocol accurately reflects the complexities of real-world daily life (French & Sutton, 2010). However, an extensive body of evidence regarding research participation effects suggests that the very act of observation may distort the phenomena under investigation, thereby threatening the ecological validity of study outcomes. Historically, such shifts in participants´ behavior have been broadly categorized under the “Hawthorne effect” (Adair, 1984), traditionally defined as behavioral modifications resulting from the awareness of being studied. Nevertheless, contemporary systematic reviews argue that this nomenclature is perhaps too reductive, subsuming a heterogeneous array of participation effects that require rigorous disentanglement to truly understand their underlying causal mechanisms (McCambridge et al., 2014).

The Dual-Nature of Self-Monitoring

The conceptual boundary between measurement and intervention is notably blurred in the context of self-monitoring. In the behavioral sciences, self-monitoring is frequently employed as a deliberate Behavior Change Technique (BCT; Abraham & Michie, 2008; Michie et al., 2013) within complex interventions (e.g., Middelweerd et al., 2014; Williams & French, 2011). Formally defined as the act of a person „***observing*** or ***recording*** their own behavior with the knowledge that it is part of a change strategy” (Michie et al., 2015), this technique serves to systematically track action to enhance self-awareness and self-regulation (Burke et al., 2011). Within this interventionist framework, measurement reactivity (MR) is not a methodological artifact to be suppressed but rather a primary therapeutic mechanism, thus, reactivity to the measurement is the desired outcome, provided the resultant behavioral shifts align with the targeted health goals.

Conversely, in non-interventional contexts such as baseline assessments in randomized controlled trials (RCTs) or large-scale epidemiological surveys, MR effects represent a significant threat to internal and external validity (Korotitsch & Nelson-Gray, 1999). This assessment-driven reactivity can be profoundly detrimental to public health and clinical research, potentially introducing biases that lead to false-positive results (Miles et al., 2018; French & Sutton, 2010; French et al., 2021). For instance, the mere act of responding to queries may prompt cognitive elaboration, where participants reflect more deeply or in novel ways on their habits, thereby catalyzing behavior change independent of the active intervention components (McCambridge et al., 2014). Such “noise” in the data may lead to the premature scaling of ineffective interventions, incurring substantial economic costs and delaying the implementation of truly efficacious strategies.

Alternatively, assessment tools may introduce a systematic “ceiling” or “floor” effect that masks genuine intervention impact. A critical example occurs when a self-monitoring device (e.g., a wearable tracker) used as the intervention tool is also utilized to establish baseline data across all study arms. This may induce a “pre-treatment” reactivity that minimizes the detectable difference between groups, leading to the erroneous conclusion that an intervention is unsuccessful (Miles et al., 2018). In digital health contexts where the user interface itself can serve as a constant prompt, MR and related mechanisms are potentially potent. In this dissertation, the reference setting is the evaluative research context where the primary objective remains the control, minimization or total elimination of these reactive artifacts to ensure the integrity of the scientific evidence base.

Measurement Reactivity

Central to this discourse is the concept of *measurement reactivity*: the phenomenon wherein a participant's behaviors, cognitions, or affective states are altered specifically due to the process of being measured (French & Sutton, 2010). This is most frequently operationalized as the question-behavior effect (Sprott et al., 2006) or mere-measurement effect (Morwitz et al., 1993) where the simple administration of an inquiry prompts a shift in subsequent action. Reactivity is thought to be induced at the beginning of a study protocol, when there is a heightened awareness of one's behaviors, cognitions, or beliefs (Barta et al., 2012). Individuals may experience self-consciousness or evaluation anxiety, present themselves in a favorable light or in line with a positive self-image, or increase effort necessary to reflect on and rate constructs of interest, ultimately resulting in modified behavior, cognitions, or beliefs (Barta et al., 2012). Nevertheless, research suggests that MR depletes and typically resolves after the first 1–2 days of a study protocol, returning to "normal" levels thereafter (French & Sutton, 2010). However, since estimates of behavior, cognitions, and beliefs are often aggregated over this period, MR may bias these estimates by skewing them meaningfully upward or downward (French et al., 2021). Meta-analyses and systematic reviews have quantified these effects across diverse health domains, consistently identifying small to moderate effect sizes, with Cohen's d ranging from 0.06 to 0.28 (McCambridge & Kypri, 2011; Miles et al., 2020; Rodrigues et al., 2015; Spangenberg et al., 2016; Wilding et al., 2016; Wood et al., 2016).

In the current landscape of health behavior research, the rise of intensive longitudinal designs, such as Ecological Momentary Assessment (EMA), has added complexity to this issue. EMA is a real-time data capture methodology in which individuals are intensively and repeatedly assessed within days on some phenomenon of interest (Stone & Shiffman, 1994; Shiffman et al., 2008). EMA protocols can include single items or brief assessments related to the antecedents of the phenomenon of interest (Stone& Shiffman, 1994). Because data from EMA studies are date- and time-stamped, it facilitates pairing of data from other sources such as for example, device-based measures of physical activity (PA) and sedentary behavior (SB).

Empirical support for the presence of MR in EMA is robust; for instance, studies of movement-related behaviors have operationalized MR in many ways. Many researchers use an experimental design and cover stories to conceal the purpose of the PA or SB monitoring device (Clemes & Parker, 2009); others sealed the display of devices, to prevent feedback from self-monitoring (Clemes & Deans, 2012). Many studies use observational designs, in which participants' behavior is tracked for several days. Reactivity is usually defined in one of two ways: (a) as a linear trend, assuming desired behaviors (e.g., PA) would gradually decrease as the study progresses, while undesired behaviors (e.g., SB) would gradually increase or (b) as a "drop-off" after initial days or observations and then consistency (Maher et al., 2024).

In addition to behavior, MR can affect motivational constructs (Barta et al., 2012). For instance, smokers completing an EMA protocol before the quit attempt had higher levels of self-efficacy immediately prior to the quit attempt, compared with smokers who completed the protocol after the quit attempt or who did not complete any EMA protocol, suggesting that the act of assessing motivational constructs may impact motivation itself (Rowan et al., 2007). University students also showed decrease in their positive affect and emotional awareness in a 14-day experience sampling protocol (ESM), which specifically investigated under which circumstances ESM induces reactive effects (Eisele et al., 2023). Whereas, soliciting intentions regarding blood donation has been shown to significantly augment actual donation rates (Godin et al., 2008), while prompts concerning the anticipated regret of non-attendance in cervical cancer screenings successfully increased clinical participation (Sandberg & Conner, 2009).

In comparison to movement-related or other consumption-related behaviors (i.e., alcohol consumption, smoking), there have been very little investigations of potential reactivity in dietary intake (König et al., 2022). Thus the (non) presence of MR, its mechanisms or related moderators in the context of dietary intake and ambulatory assessment, are yet to be investigated.

Digital and Ambulatory Assessment

To mitigate the inherent limitations of retrospective self-reports such as questionnaire data, researchers have increasingly pivoted toward objective behavioral assessment via digital technologies. This shift is characterized by the deployment of wearable sensors, such as pedometers and accelerometers (e.g., Presseau et al., 2013; Sasaki et al., 2016), and the utilization of smartphone-embedded sensors (Miller, 2012) to quantify for example PA. Similar advancements are evident in clinical settings, where electronic medication monitoring systems provide high-fidelity data on adherence (Sutton et al., 2014).

The primary advantage of these digital tools lies in their ability to record behavior in situ, thereby bypassing the cognitive burden of recall and significantly reducing retrospective recall bias (Shiffman et al., 2008). Theoretically, this proximity to the behavioral event should enhance the ecological validity of the collected data. However, a critical methodological question persists: are these ostensibly "objective" digital measures themselves susceptible to MR? While digital tools remove the subjective bias of *reporting*, they do not necessarily remove the psychological impact of *monitoring*.

Furthermore, the adoption of ambulatory assessment (AA) has seen a significant upward trajectory (Hamaker & Wichers, 2017). Alternately conceptualized as daily diary methods, experience sampling, or EMA, AA represents a suite of methodologies designed to capture individuals' lived experiences including ongoing behavior, subjective experience, physiology, and environmental contexts within naturalistic and unconstrained settings

(Fahrenberg et al., 2007; Mehl & Conner, 2012). The proliferation of AA is driven by several established methodological advantages: the attainment of high ecological validity (e.g., Trull & Ebner-Priemer, 2013), the capacity to capture complex within-person dynamics (Hamaker & Wichers, 2017), and the substantial mitigation of retrospective recall bias (Conner & Feldman Barrett, 2012).

Despite these strengths, a critical methodological question remains unresolved: whether, or under what specific conditions, these repeated assessments inadvertently induce MR, altering so the very constructs they intend to measure (Ottenstein et al., 2024). While reactivity can occur during isolated, one-time assessments (Webb et al., 2000), it is particularly salient in AA designs where participants engage with the same inquiries repeatedly. The cognitive process of responding to an AA questionnaire reading items, reflecting on their meaning, monitoring internal states, and selecting responses (Tourangeau et al., 2000) naturally heightens the participant's attention toward the measured constructs. For example, smartphone-based EMA facilitates the granular assessment of diverse health behaviors, including dietary patterns (König et al., 2021), by enabling the objective capture of food choices for instance, through time-stamped photographic evidence of meals before and after consumption (König & Renner, 2018, 2019; Martin et al., 2012).

This intensified self-monitoring may, in turn, alter the participant's behaviors, cognitions, or emotions throughout the study period or beyond (Barta et al., 2012). The repetitive nature of AA protocols may further exacerbate this effect. As participants become familiar with the sampling schedule and item content, they may modify their responses or behaviors in anticipation of subsequent prompts.

These findings present a “validity paradox” for researchers: while digital tools provide more precise measurements than traditional questionnaires, the heightened transparency and constant feedback loops inherent in digital devices may inadvertently amplify reactive effects.

This suggests that objective data may provide an accurate measurement of a behavior that has already been fundamentally altered by the assessment process itself. Consequently, systematic investigation into MR mechanisms, how it is moderated by specific constructs, participant characteristics, and study design parameters is essential for the continued advancement of digital ambulatory assessment methodologies and ensuring the accuracy of behavioral data in both clinical trials and public health surveillance.

Research Gap: Measurement Reactivity in Dietary Assessment

While the impact of digital monitoring on PA is well quantified, empirical data regarding reactivity in other health behaviors remain sparse. At the date when this dissertation was conceptualized, no studies had specifically investigated MR within the context of digital dietary assessment. Since then, only one study was published testing potential reactivity effects to dietary EMA. In this study, Whitton and colleagues aimed to identify demographic and psychosocial correlates of measurement error and reactivity bias, among adults who used a 4-day image-based mobile food record through which energy intake (EI) was measured. While the total sample showed a modest 3% daily decrease in EI, some participants who were labelled Reactive Reporters showed a substantial 17% decrease in EI per day of recording. Overall, Whitton et al. suggest that even "low-burden" image-based tracking methods are susceptible to significant reactivity (Whitton et al., 2023). This represents a significant gap in the literature, particularly as large-scale research initiatives in nutrition research increasingly rely on smartphone-based tools to shape future clinical practice and research agendas.

The evidence for dietary reactivity is currently contradictory. On one hand, meta-analytic data on the question-behavior effect suggests only a small, non-significant effect for dietary behaviors (SMD = 0.21; Miles et al., 2020). On the other hand, the inclusion of self-monitoring in healthy eating interventions consistently yields larger effect sizes (Michie et al., 2009), suggesting that some degree of reactivity is likely occurring but remains poorly understood outside of an interventionist context. Furthermore, the temporal dynamics of this effect are unclear. While PA reactivity may be short-lived (Clemes & Deans, 2012), evidence from medication adherence suggests that stabilization may take upwards of six weeks (Cook et al., 2012). Determining the "run-in" period necessary for dietary assessment is therefore essential for methodological rigor.

Beyond establishing *if* reactivity occurs, there is a pressing need to understand *how* and *why* it occurs (French & Sutton, 2010). Traditional explanations center on social desirability and the awareness of being observed (McCambridge et al., 2014) or increased attitude accessibility (Morwitz & Fitzsimons, 2004). However, evidence from pedometer research suggests a different mechanism: the degree of participants´ interaction with the device. Clemes and Parker (2009) demonstrated that step counts increased significantly only when participants were required to actively log their data in a diary, rather than just wearing a sealed or unsealed device. This suggests that the requirement to interact with the recording tool may be the primary driver of behavioral change.

From a theoretical perspective, interaction with an assessment tool may induce reactivity by heightening awareness and prompting reflection on the target behavior (Barta et al., 2012). According to self-regulation theory (Carver & Scheier, 1982; Kanfer & Gaelick-Buys, 1991), if this reflection reveals a discrepancy between current and desired behavior, self-regulatory processes are triggered to align the two. Consequently, the degree of AIP (the active reflection on the information that the participants have to provide while recording their behavior) or the cognitive load and time required for data entry may moderate the magnitude of reactivity. This hypothesis is particularly relevant for smartphone-based dietary assessment, which varies significantly in participant involvement (König et al., 2018; König et al., 2021). Detailed dietary records requiring manual database entry are time-consuming and likely prompt deep reflection (König et al., 2021), whereas image-based "photo-logs" (e.g., Boushey et al., 2017; König & Renner, 2018) may be completed in seconds with minimal cognitive effort.

This dissertation aims to address these gaps in the literature by examining the presence of MR in digital dietary intake, and the possible moderators of this participation effect in an ecological ambulatory assessment setting.

1.2. Aim and outline of present dissertation

The four research papers part of this dissertation, summarized herein build on each-other aiming to address a relevant query in modern health psychology: the trade-off between the high ecological validity of digital in-the-moment assessments and the potential for these assessments to inadvertently alter the behaviors they aim to measure. The dissertation's overarching aim is to determine whether smartphone-based assessments of dietary intake are prone to reactive, and to identify assessment tool-related and person-related factors that might moderate this effect.

Firstly, a systematic review and meta-analysis was conducted (chapter 2 - *A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour*). Through the systematic review, the aim was to (1) consolidate the fragmented literature regarding reactivity to digital in-the-moment measurements, providing thus an evidence-based foundation for the dissertation. Additionally, it was aimed to (2) quantify the effect size of MR across various health behaviors (e.g., PA, diet, alcohol consumption) and to identify moderators that might explain the variability in findings across previous studies. The suggestions retrieved from this systematic review and meta-analysis, directly informed the design of studies 2, 3 and 4 of this dissertation.

In the second study (chapter 3 - *The same, only different: Smartphone-based dietary Ecological Momentary Assessment tools vary in complexity, usability and active information* *processing*) we shift the focus to the assessment tool itself (smartphone-based dietary intake apps), investigating the variability of the design and complexity of EMA tools, and the perceived potential influence of AIP (and thus reactivity).

As suggested by the literature and supported by the findings of the first study, reactivity varies across studies also depending on the tool used, and it is driven by heightened self-monitoring, we assumed that also the cognitive effort required to use an EMA tool might influence the degree of reactivity. We hypothesized that smartphone-based dietary EMA tools ranging from simple checklists, photo-based assessment, to complex food-entry databases, differ significantly in (1) their inducted AIP (*the extent to which the tool induced participants to reflect on their dietary choices during the reporting process*) requirements, (2) in the complexity, and (3) usability. Moreover, it was hypothesized that (4) these differences might influence eating-related cognitions, intentions and behaviors.

This second study employed a comparative, between-subjects design where participants were randomly exposed to one of eight different types of smartphone-based dietary EMA interfaces (EMA protocols) in the form of mock-ups and series of screenshots in combination to the detailed written description of the diet intake recording process. The EMA protocols differed in the type and number of tracking features combined e.g., photo-based assessment; food database; detailed description; portion size estimation. Additionally, the perceived usability (using the User Experience Questionnaire), previous experiences with nutrition apps and the affinity to technology were assessed as well. Since the study was based on a hypothetical scenario (participants were shown mock-up versions of the smartphone-based app), it guided the design of the EMA-smartphone-app tested in real life in study four.

Whereas the third study (chapter 4 - *Easy or difficult? Investigating perceived ease of changing eating and physical activity behaviours*) focuses on participants' perceptions of behavior. The aim was to explore potential differences in the perceived ease of change (a key psychological moderator to MR as identified in the first study) of 21 specific aspects of behavior (13 eating-related and 8 PA-related) and (2) potential moderators. In two preregistered online studies (in Germany and USA), young adults predominantly without and older adults predominantly with chronic conditions indicated the perceived ease of changing 21 aspects of eating and PA, medical history, social comparison, prior behavior change attempts and current behavior.

This study refined the "target" of our reactivity research by suggesting that in study four, we include a range of dietary behaviors with varying perceived ease of change scores to see if reactivity is indeed moderated by the participant's belief in their own plasticity. Consequently, this study justifies the selection of specific outcome variables in the final empirical test (Study four – chapter 5), ensuring that we test reactivity also across a spectrum of "easy" and "difficult" behaviors rather than treating eating-related behavior as a monolithic entity.

Lastly, the fourth study (chapter 5 - *Measurement reactivity in smartphone-based dietary ecological momentary assessment*) serves as the empirical capstone, utilizing the insights from the previous three studies to conduct a well-powered, smartphone-based study to quantify reactivity in dietary EMA. This study investigated (1) whether MR occurred in an eight-day event-based smartphone-based dietary assessment. It further tested (2) which aspects of eating behavior were most susceptible, and (3) the moderating roles of AIP, previous experience with diet-tracking apps, and the intention to change eating behavior were explored. Generally, healthy behaviors were expected to decline and unhealthy behaviors to increase during the study period. Eighty-two participants (within two data collection waves, namely in Germany and Austria) recorded all meals and snacks by taking pictures of their meals/snacks and by providing text descriptions of the meals/snacks. This final study served as the empirical “capstone” of the dissertation closing the loop appended by the previous studies.

All four studies, part of this dissertation, follow open science and data sharing principles and have been pre-registered in the OSF platform. All details of the pre-registrations and data availability are found in each of the individual papers. Studies two, three, and four received ethical approval from the Bayreuth University Ethics Committee.

2. A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour

Laura M König [1, 2], Anila Allmeta [1], Nora Christlein [1], Miranda Van Emmenis [2], Stephen Sutton [2]

[1]Faculty of Life Sciences: Food, Nutrition and Health, University of Bayreuth, Kulmbach, Germany

[2]Behavioural Science Group, Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

This chapter has been published in:

König, L. M., **Allmeta, A.,** Christlein, N., Van Emmenis, M., & Sutton, S. (2022). A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour. *Health Psychology Review*, *16*(4), 551–575. https://doi.org/10.1080/17437199.2022.2047096

Contribution:

SS and LK conceived of the study and developed the research question. LK developed the search strategy with input from SS, **AA** and MVE. Searches were conducted by LK; screening was conducted by LK, **AA**, MVE and NC; data was extracted by LK and NC; study quality was appraised by **AA**, LK and MVE. Data was analyzed and interpreted, and the manuscript was written by LK with input from all authors. All authors approved the final version of the manuscript.

2.1. Abstract

Objectives: Self-report measures of health behaviour have several limitations including measurement reactivity, i.e., changes in people′s behaviour, cognitions or emotions due to taking part in research. This systematic review investigates whether digital in-the-moment measures induce reactivity to a similar extent and why it occurs.

Methods: Four databases were searched in December 2020. All observational or experimental studies investigating reactivity to digital in-the-moment measurement of a range of health behaviors were included if they were published in English in 2008 or later.

Results: Of the 11,723 records initially screened, 30 publications reporting on 31 studies were included in the qualitative synthesis/ 7 studies in the quantitative synthesis. Eighty-one percent of studies focused on reactivity to the measurement of physical activity indicators; small but meaningful pooled effects were found (Cohen′s ds: 0.27–0.30). Only a small number of studies included other behaviors, yielding mixed results.

Conclusion: Digital in-the-moment measurement of behaviour thus may be as prone to reactivity as self-reports in questionnaires. Measurement reactivity may be amplified by (1) ease of changing the behaviour (2) awareness of being measured and social desirability, and (3) resolving discrepancies between actual and desired behaviour through self-regulation.

2.2. Introduction

In behavioural science, including psychology, self-report measures of behaviour are ubiquitous (Baumeister et al., 2007). Self-report measures provide researchers with information on people's behaviour at a comparably low cost, e.g., when data is collected via (online) questionnaires. However, they suffer from several shortcomings. Self-report assessments are usually retrospective one-time measures that require participants to average across many different occasions, inducing a recall bias (Coughlin, 1990). For instance, Food Frequency Questionnaires typically ask participants to indicate the frequency of consuming certain categories of food within the past months or year (e.g., Winkler & Döring, 1998; see also Thompson and Subar, 2017, for a discussion). Furthermore, responses may be biased because of social desirability, i.e., the participant answering in a way that they feel will satisfy the researcher instead of reporting their actual behaviour or emotions (Grimm, 2010). Finally, self-report measures such as questionnaires have been shown to be prone to measurement reactivity (Barta et al., 2012; Ram et al., 2017), i.e., changes in people's behaviour, emotions or cognitions due to being measured as part of a research project (French & Sutton, 2010). Since measurement reactivity has mainly been studied in the context of questionnaire-based studies, it is also referred to as the question-behaviour effect (Sprott et al., 2006) or mere-measurement effect (Morwitz & Fitzsimons, 2004; Zajonc, 1968). Several recent systematic reviews and meta-analyses have summarized the evidence on the impact of completing a (pen-and-paper or online) questionnaire on a range of behaviours. They consistently report small but significant effect sizes of Cohen's d ranging from 0.06 to 0.28 (McCambridge & Kypri, 2011; Miles et al., 2020; Rodrigues et al., 2015; Spangenberg et al., 2016; Wilding et al., 2016; Wood et al., 2016). It is thus questioned whether self-report measurements allow researchers to draw sufficiently valid conclusions about human behaviour as well as its determinants and consequences (French & Sutton, 2010).

To overcome these limitations, it is often recommended to use objective measures of behaviour or to reduce the time span between the behaviour occurring and it being assessed by using Ecological Momentary Assessment (Shiffman et al., 2008). Because of recent technological developments and reduced costs, the use of digital measurement devices is becoming increasingly popular especially in health behaviour research including health psychology (Müller et al., 2018). For instance, health-related behaviours such as physical activity and sedentary behaviour are increasingly tracked using wearable devices (Reichert et al., 2020), while consumption behaviours such as eating are increasingly studied using smartphones (König et al., 2021) and body-worn sensors (Bell et al., 2020). These methods provide detailed insights into health behaviours in daily life and allow individuals to measure behaviour when it occurs, thereby reducing recall bias and increasing the validity of the collected data (Shiffman et al., 2008). Accordingly, measuring behaviour immediately when it occurs in daily life using digital devices is assumed to have fewer methodological shortcomings than self-reports assessed via questionnaires. It is unclear, however, whether the validity of behavioural data that is recorded with digital devices is also influenced by research participation effects such as measurement reactivity (French & Sutton, 2010). For example, when using digital measurement devices, people are typically aware of the study context. Accordingly, it could be hypothesized that also digital measurement of behaviour suffers from measurement reactivity.

Measurement reactivity may be especially challenging in the behavioural and health sciences, where digital assessments of behaviour are increasingly used to study determinants of behaviour and to evaluate the effectiveness of interventions (e.g., Degroote et al., 2020a; König et al., 2021). In contrast to self-monitoring, which is used as a Behaviour Change

Technique to deliberately induce changes in behaviour through recording (Michie et al., 2015), measurement reactivity is usually undesired since it distorts the study findings. For example, recording a behaviour might induce reflecting on the behaviour, which might increase the likelihood of behaviour change independent of intervention components (Barta et al., 2012; McCambridge et al., 2014). This, in turn, may lead to ineffective interventions being implemented on a larger scale, so creating unnecessary costs and preventing more effective interventions from being implemented. On the other hand, assessment tools might introduce systematic bias that conceals true intervention effects, e.g., when a self-monitoring device used for the intervention is also used for baseline measurements in intervention and control groups. The tool may impact behaviour in all participants and thus lead to the erroneous conclusion that the intervention was unsuccessful (Miles et al., 2018; see also French et al., 2021b, for a discussion). Taking potential measurement reactivity into account when analysing behavioural data by introducing it as a model parameter may be crucial; however, this is only possible if an estimate of the effect is available (Bendtsen & McCambridge, 2021; French et al., 2021b). The present systematic review updates an earlier rapid review (see French et al., 2021a; French et al., 2021b, for summaries) to synthesize the evidence on and the magnitude of reactivity to digital in the- moment measurement of health behaviour, to guide future research activities on measurement reactivity and extend existing guidance on reducing measurement reactivity in behavioural and clinical research (French et al., 2021b).

2.3. Methods

A protocol was developed following the PRISMA 2009 guidelines (Moher et al., 2009) and registered on PROSPERO (registration number CRD42021221933) prior to conducting this systematic review. This report follows the updated PRISMA 2020 guidelines

(Page et al., 2021). Raw data and analysis scripts are available on the project's Open Science Framework page: https://osf.io/qnmvj/.

Inclusion and exclusion criteria

Any observational or experimental study using a digital (e.g., smartphone app, pedometer, medication event monitoring system) tool to assess behavioural data repeatedly in daily life was included. Study protocols, systematic reviews and meta-analyses were excluded. Studies were also excluded if assessments were not digital (e.g., paper diaries). Behaviours were restricted to the following health behaviours which are among the leading health risk factors (GBD, 2019 Viewpoint Collaborators, 2020): alcohol consumption, dental care, diet, medication adherence, physical activity, sedentary behaviour, smoking. Studies were included if they aimed to investigate reactivity to the measurement, i.e., changes in behaviour due to being measured as part of a research project (French & Sutton, 2010), comparing records either between different conditions (e.g., comparing different devices) or within participants across conditions or over time. Self-monitoring interventions or interventions providing feedback that aimed to change participants′ behaviour through tracking were excluded. There were no restrictions regarding the participants′ age or health status. Studies had to be published in peer-reviewed journals and written in English; accordingly, studies published in any other language as well as theses and preprints were excluded.

Search strategy

An electronic search strategy was developed based on the inclusion criteria. It included keywords related to the different health behaviours and tools for their assessment (e.g., physical activity, pedomet*) and measurement/ assessment reactivity. The strategy was initially developed for Pubmed (incl. MEDLINE) and adapted for PsycINFO, Embase and

Web of Science Core Collection (see Appendix A for the strategies developed for the different databases). All databases were searched from 2008 to 1st December 2020, when the search was conducted. The publication date was restricted to 2008 onwards since research in digital assessment of health behaviour has accelerated since then through the development of smartphones, and the vast majority of papers in this field have been published in the 2010s (Müller et al., 2018). Reference lists of all eligible papers were hand searched and forward citation tracking using Google Scholar was conducted to identify further eligible publications. Moreover, emails were sent to the members' mailing lists of the European Health Psychology Society, the German Psychological Society and the British Psychological Society to identify further eligible work.

Study selection

All records retrieved from the database searches were imported into Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia; available at www.covidence.org). Duplicates were removed before titles and abstracts were screened independently by two authors, coding articles as provisionally eligible or excluded according to the inclusion and exclusion criteria. Disagreements were resolved by discussion. Afterwards, full texts were screened independently by two authors and coded as eligible or excluded. Again, disagreements were resolved by discussion. A PRISMA flow diagram (Page et al., 2021) documents the flow of records (see Figure 1).

Data extraction and synthesis

For all eligible studies, two authors independently extracted study characteristics relevant to categorising and describing the studies (e.g., target behaviour[s], study design, description of the digital assessment tool[s], description of condition[s], moderators) and outcomes (effect size[s] or relevant indices needed for calculating the effect size, overall conclusion regarding the presence of measurement reactivity) in a form that was developed based on a previous rapid review (French et al., 2021a; see Appendix B for the full list of information extracted). Discrepancies were identified and resolved by discussion. Study authors were contacted to obtain key unpublished outcome data. Data on all studies were synthesised narratively.

In addition, a meta-analysis was conducted for experimental studies on reactivity to measuring all physical activity indicators combined and for steps only, using MAJOR – Meta-Analysis for Jamovi Version 1.2.0 (Hamilton, 2018) with Jamovi 1.6.23 (The jamovi project. jamovi [Version 1.6.23]. 2021). Random effects models were computed to calculate pooled effect sizes. Since all experimental studies compared continuous data collected from at least two groups, Cohen's d is reported. If available, means and standard deviations were extracted from the studies to calculate Cohen's d (Cohen, 1992) following the recommendations outlined in Borenstein et al. (2009) and The Cochrane Collaboration (2011b). If the information provided in the publication was not sufficient to calculate Cohen's d, authors were contacted and asked to provide the effect size or raw data. For three studies, no information about effect sizes could be obtained (Clemes & Deans, 2012; Foote et al., 2017; McCarthy et al., 2015). Heterogeneity of effect sizes was evaluated using I2 as recommended by Higgins et al. (2003). Due to a small number of studies for individual combinations of study manipulations, it was not possible to calculate separate meta-analyses. In addition, individual effect sizes for all experimental studies are reported in Appendix C.

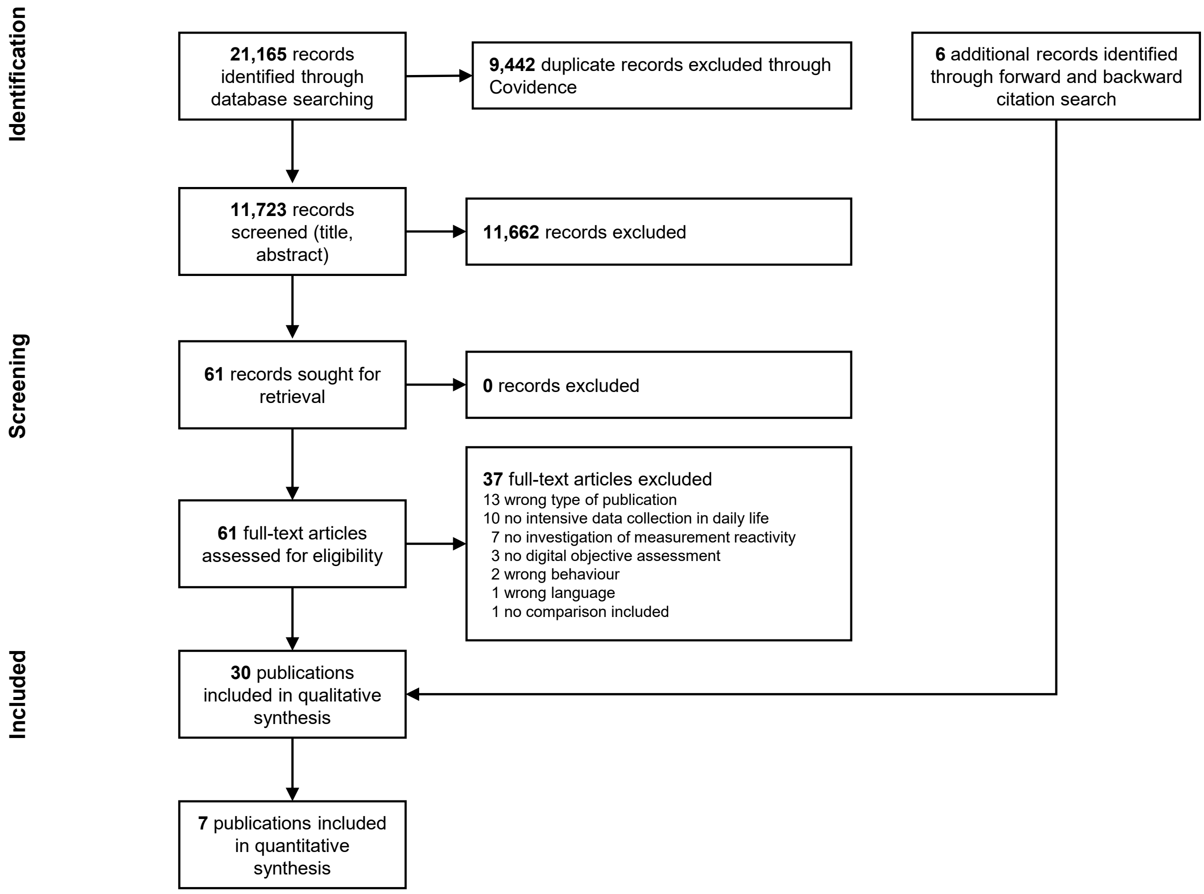


Figure 1. PRISMA flow chart.

Risk of bias

For experimental studies with randomised group allocation, risk of bias was assessed using the Cochrane Risk of Bias 2.0 tool (Sterne et al., 2019). Studies using a within-subjects design were appraised using the checklist for cross-over studies by Ding et al. (2015). For observational studies, risk of bias was assessed using the JBI Checklist for Analytical Cross Sectional Studies (Moola et al., 2020). Furthermore, a potential publication bias was investigated using funnel plots and Egger's test for funnel plot asymmetry (Egger et al., 1997). Funnel plots were adjusted using the trim and fill method (Duval & Tweedie, 2000) using metafor 3.0-2 (Viechtbauer, 2010) in R Studio 1.1.456/ R version 4.0.3.

Deviations from the protocol

Since truly objective digital assessment of consumption behaviours in daily life including smoking, alcohol consumption, and food intake is still in its early stages (Höchsmann & Martin, 2020; König et al., 2021), it was decided prior to screening titles and abstracts that Ecological Momentary Assessment (EMA) (Shiffman et al., 2008) of consumption behaviours would be considered as long as participants were asked to record the occasion immediately before, during or after consumption, preferably using objective markers such as a photo (König et al., 2021) which minimises recall bias (Shiffman et al., 2008).

2.4. Results

Literature search

A total of 11,723 individual records were screened. After 11,662 were excluded when screening titles and abstracts, 61 full texts were screened for eligibility. An additional 6 records were identified through forward and backward citation searches. A total of 30 publications reporting on 31 studies were included (see Figure 1 for the flow of records). Two of these studies overlap in the reported data: Ullrich et al. (2021) extends the data reported in Baumann et al. (2018) by adding data from a second point of measurement.

Study and sample characteristics

The 30 publications were published between 2008 and 2021. Most publications stemmed from the US (23%, n = 7) (Davis & Loprinzi, 2016; Foote et al., 2017; Ling & King, 2015; McCarthy et al., 2015; Prewitt et al., 2013; Yang et al., 2015; Zhu & Haegele, 2019), followed by the UK (17%, n = 5) (Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009; Ho et al., 2013; Sutton et al., 2014) and Australia (10%, n = 3) (Poulton et al., 2019; Scott et al., 2014; Tinlin et al., 2018). See Table 1 for a summary of the study characteristics.

The majority of the 31 included studies focused on different aspects of physical activity (81%, n = 25) including steps/ walking (n = 15) (Clemes et al., 2007; Clemes &

Deans, 2012; Clemes & Parker, 2009; Craig et al., 2010; Haegele et al., 2020; Hilgenkamp et al., 2012; Klenk et al., 2019; Ling et al., 2011; Ling & King, 2015; Motl & Dlugonski, 2011; Motl et al., 2012; Prewitt et al., 2013; Scott et al., 2014; Tinlin et al., 2018), moderate to vigorous physical activity (MVPA; n = 9) (Baumann et al., 2018; Davis & Loprinzi, 2016; Haegele et al., 2020; Ho et al., 2013; Tinlin et al., 2018; Ullrich et al., 2021; Vanhelst et al., 2017; Zhu et al., 2020; Zhu & Haegele, 2019), light physical activity (LPA; n = 6) (Baumann et al., 2018; Haegele et al., 2020; Tinlin et al., 2018; Ullrich et al., 2021; Vanhelst et al., 2017; Zhu et al., 2020) and activity counts (n = 7) (Davis & Loprinzi, 2016; Dössegger et al., 2014; Foley et al., 2011; Foote et al., 2017; Ho et al., 2013; Motl & Dlugonski, 2011; Vanhelst et al., 2017). Four studies also investigated reactivity to measurement of sedentary behaviour (13%) (Baumann et al., 2018; Tinlin et al., 2018; Ullrich et al., 2021; Vanhelst et al., 2017). The remaining six studies focused on consumption behaviours such as alcohol consumption (10%, n = 3) (Labhart et al., 2020; Poulton et al., 2019; Yang et al., 2015) and smoking (3%, n = 1) (McCarthy et al., 2015) as well as medication adherence (7%, n = 2) (Cook et al., 2012; Sutton et al., 2014). One study also assessed the number of non-alcoholic drinks (3%) (Labhart et al., 2020).

The majority of studies focused on adults (55%, n = 17) (Baumann et al., 2018; Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009; Cook et al., 2012; Hilgenkamp et al., 2012; Klenk et al., 2019; Labhart et al., 2020; McCarthy et al., 2015; Motl et al., 2012; Motl & Dlugonski, 2011; Poulton et al., 2019; Sutton et al., 2014; Tinlin et al., 2018; Ullrich et al., 2021; Yang et al., 2015), while twelve studies (39%) included children or adolescents in various age ranges between 3 and 18 years (Craig et al., 2010; Dössegger et al., 2014; Foley et al., 2011; Foote et al., 2017; Haegele et al., 2020; Ho et al., 2013; Ling et al., 2011; Ling & King, 2015; Prewitt et al., 2013; Scott et al., 2014; Vanhelst et al., 2017; Zhu et al., 2020). Two studies (7%) specifically compared children, adolescents and adults (Davis & Loprinzi, 2016; Zhu & Haegele, 2019).

Table 1. Characteristics of the included studies.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
Baumann et al. (2018)[b]	Germany	PA (LPA, MVPA), SB (sitting time)	Accelerometer (ActiveGraph GT3X+)	7	adults	participants of a cardio-preventive health examination programme	160	Observational, within-subjects	1		Season (spring, summer, winter), first day of measurement (weekday vs weekend day)	(+/-) There is reactivity to measurement of LPA and SB. The effect is not significant for MVPA. Season moderates reactivity to measurement of LPA and SB with a steeper increase in SB and steeper decline in LPA in summer vs spring or winter. First day of measurement moderates

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
												reactivity to measurement of MVPA with stronger decline if measurement started on a weekday vs a weekend day.
Clemes et al. (2008)	UK	PA (steps)	pedometer (New-Lifestyles NL-2000)	14	adults		50	Experimental, within-subjects	2	covert: pedometer concealed as a body posture monitor; unsealed + diary: use of pedometer known, in addition participant were asked to copy the daily step counts into a diary		(+) There is reactivity to measurement of steps.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
Clemes and Parker (2009)	UK	PA (steps)	pedometer (New Lifestyles NL-1000)	28	adults		63	Experimental, within-subjects	4	covert: pedometer concealed as a body posture monitor; sealed: use of pedometer known, but display not visible; unsealed: use of pedometer known, display visible; diary: use of pedometer known, display visible, asked to copy daily	gender	(+) There is reactivity to measurement of steps; effect is most pronounced when wearing an unsealed pedometer and recording step counts. Gender does not moderate the effect.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
										step counts into diary		
Clemes and Deans (2012)	UK	PA (steps)	pedometer (New Lifestyles NL-1000)	21	adults		90	Experimental, within-subjects; also studies change in behaviour across time (study weeks)	2	covert: pedometer was concealed as a body posture monitor; overt: use of pedometer was announced, participants were asked to copy step counts into diary	BMI group, employment status (staff/ student), sex	(+) There is reactivity to measurement of steps; effect washes out after one week. BMI group, employment status and sex do not moderate the effect.
Cook et al. (2012)	NA	Medication adherence (% of prescribed doses taken)	Medication Event Monitoring System (MEMS)	84	adults	Patients with glaucoma	45	Observational, within-subjects	1			(+) There is reactivity to measurement of medication adherence; effect washes out after 5 weeks.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
Craig et al. (2010)	Canada	PA (steps)	pedometer (SW-200)	7	Children and adolescents (5 to 19 years)		11477	Observational, within-subjects	1			(-) There is no reactivity to measurement of steps
Davis and Loprinzi (2016)	US	PA (activity counts, MVPA)	accelerometer (ActiGraph 7164)	7	Children (6 to 11 years), adolescents (12-17 years), adults (18 to 85 years)		674	Observational, within-subjects	1		Age, gender, first day of monitoring (day of the week; weekday vs weekend day)	(-) There is not reactivity to measurement of MVPA or activity counts. There was some evidence that reactivity to measurement of activity counts was stronger in adults if the first day was a Monday.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
Dössegger et al. (2014)	Switzerland	PA (activity counts)	accelerometer (ActiGraph models 7164, GT1M, GT3X)	7	Children, adolescents (3 years and older)		2081	Observational, within-subjects, data pooled from 8 studies	1		Age group (3-6 years, 7-11 years, ≥12 years), first day of monitoring (day of the week)	(+) There is reactivity to measurement of activity counts. The first day of monitoring moderated the effect; the effect was stronger for participants who started to monitor on a Wednesday compared to Sunday. Also age moderated the effect; measurement reactivity was found starting from the age of 7.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
Foley et al. (2011)	NA	PA (activity counts)	pedometer (Walk4Life 2525), accelerometer (Actiwatch)	20	Children (7-11 years)		32	Experimental, within-subjects; also studies changes in behaviour across time (within summer school blocks)	2	Treatment condition: participant wore both an accelerometer which was concealed as a watch to measure the time, and a pedometer which measured steps No treatment condition: Participants only wore the concealed accelerometer	Time of day (according to summer camp schedule)	(+/-) There is some evidence for reactivity to measurement of activity counts. This effect occurred in a warm-up period, but not for the overall assessment.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
Foote et al. (2017)	US	PA (activity counts)	accelerometer (MOVABLE MOVband3)	16	children (10-12 years)		25	Experimental, within-subjects; also studies change in behaviour across time (study weeks)	2	Sealed accelerometer: participants were unable to see the activity counts (first study week); Unsealed accelerometer: participants were able to see the activity counts on the device (weeks 2-4)	Activity in vs after school	(+) There is reactivity to measurement of activity counts, but only when wearing unsealed accelerometers: move counts were higher in first week of wearing the unsealed accelerometer vs the second week of wearing the accelerometer, while there was no significant difference between the week in which

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
												the sealed accelerometer was worn vs the first week when the unsealed accelerometer was worn. Effects for reported separately for in school and out-of-school activities, but no test of an interaction was reported.
Haegele et al. (2020)	NA	PA (LPA, MVPA, step count/ wear time ratio)	accelerometer (ActiGraph GT3X)	7	Adolescents (13-18 years)	Autisim spectrum disorder	23	Observational, within-subjects	1			(+/-) There is no statistically significant effect of reactivity to the indicators of PA, but change

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
												exceeds recommendation of 5% deviance on first or last day of measurement.
Hilgenkamp et al. (2012)	The Netherlands	PA (steps)	pedometer (Yamax Digi-Walker, NL-2000, NL-1000, NL-800)	14	adults (50 years and older)	with borderline to severe intellectual disability	135	Observational, within-subjects	1		Gender, age (below or above 65 years), level of intellectual disability, participants with/ without down syndrome	(-) There is no reactivity to measurement of steps. No moderators were identified.
Ho et al. (2013)	UK	PA (activity counts, MVPA)	accelerometer (Actiheart), pedometer (OMRON HJ-109)	4	adolescents (mean age 14.5 years)		892	Experimental, between-subjects; also studies change in behaviour	2	With pedometer: participants wore a pedometer in addition to an	Gender	(+/-) There is some evidence for reactivity to measurement of activity

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
								across time (study days)		accelerometer; Without pedometer: participants only wore an accelerometer		counts, but not for MVPA. Gender moderated the effect: viewing step counts of the pedometer was only associated with increased activity counts in girls.
Klenk et al. (2019)	Germany	PA (walking time)	Accelerometer (activPAL)	5-7	Adults (65+ years)		1333	Observational, within-subjects	1			(-) There is no reactivity to measurement of walking duration.
Labhart et al. (2020)	Switzerland	Alcohol consumption (number of alcoholic drinks), diet	smartphone app (Youth@Night, Android)	49	adults (16-25 years)		241	Observational, within-subjects	1		Commitment level (assiduous, regular, and irregular	(-) There is no reactivity to measurement of consumption of alcoholic and non-

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
		(number of non-alcoholic drinks)									participants)	alcoholic drinks. Commitment level did not moderate the effect.
Ling et al. (2011)	Hong Kong	PA (steps)	pedometer (New Lifestyles NL-800)	21	children (9-12 years)		133	Observational, within-subjects	1		Rehearsal score	(+) There is reactivity to measurement of steps. The effect was more pronounced in high rehearsers.
Ling and King (2015)	US	PA (steps)	pedometer (Yamax SW-200)	7	children (mean age 9.25 years)		126	Observational, within-subjects	1			(-) There is no reactivity to measurement of steps.
McCarthy et al. (2015)	US	Smoking (abstinence, cessation)	EMA device	28	Adults	smokers trying to quit who smoked at least 10	110	Experimental, between-subjects	2	high frequency condition: six prompts per day;		(-) There is no reactivity to measurement of smoking when

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
						cigarettes per day				low frequency condition: one prompt per day		comparing high and low frequency recording.
Motl and Dlugonski (2011)	NA	PA (activity counts, steps)	accelerometer (ActiGraph 7164)	21	Adults	Multiple sclerosis patients	18	Observational, within-subjects	1			(+) There is reactivity to measurement of steps and activity counts.
Motl et al. (2012), Study 1	NA	PA (steps)	accelerometer (first 7 days: ActiGraog 7164; second 7 days: Omron HJ-720ITC)	14	Adults	Multiple sclerosis patients	18	Observational, within-subjects	1			(+) There is reactivity to measurement of steps.
Motl et al. (2012), Study 2	NA	PA (steps)	accelerometer (first 7 days: ActiGraog 7164; second 7 days:	14	Adults	Multiple sclerosis patients	20	Observational, within-subjects	1			(+) There is reactivity to measurement of steps.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
			Omron HJ-720ITC)									
Poulton et al. (2019)	Australia	Alcohol consumption (drinks per day)	smartphone app (CNLab-A, iOS and Android)	14	adults (16+ years)		671	Observational, within-subjects	1		Hazardous drinking (AUDIT)	(+) There is reactivity to measurement of alcohol consumption. The effect was only evident in hazardous drinkers.
Prewitt et al. (2013)	US	PA (steps)	pedometer (Yamax Digi-Walker SW-200)	8	Children (4th to 6th grade)		109	Experimental, within-subjects	2	Sealed: steps counts were not visible; Unsealed: step counts were visible	Gender, grade, knowledge score quiz	(-) There is no reactivity to measurement of steps. The effect was not modulated by any of the moderators.
Scott et al. (2014)	Australia	PA (steps)	pedometer (Yamax Digi-Walker CW700), acceleromete	7	adolescents (13-14 years)		96	Experimental, between-subjects; also studies change in	3	daily sealed pedometer: step counts were recorded		(+) There is reactivity to measurement of steps.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
			r (Actiigraph GT3X+)					behaviour across time (study days)		daily by a research and a new sticker is put on the device; weekly sealed pedometer: display was sealed with a sticker; step counts were recorded by a researcher after the 7 days; unsealed: no sticker on the device		
Sutton et al. (2014)	UK	Medication adherence (taking medication	electronic medication-monitoring device (TrackCap)	56	adults	With Type 2 Diabetes	226	Observational, within-subjects; data from larger RCT	1			(+/-) There may be reactivity to measurement of medication

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
		as prescribed)						comparing digital and non-digital assessment approaches for medication adherence				adherence, but effects were small and not significant.
Tinlin et al. (2018)	Australia	PA (LPA, MVPA, standing time, stepping time, steps), SB (sitting time)	accelerometers (activPAL3, Actigraph GT3X, Sensewear)	7	adults	Stroke history	32	Observational, within-subjects	1			(+/-) There is no reactivity to measurement of sitting time, standing time, LPA or MVPA. There is reactivity to measurement of steps and stepping time.
Ullrich et al. (2021) [b]	Germany	PA (LPA, MVPA), SB (sitting time)	accelerometer (ActiGraph GT3X+)	14	adults	participants of a cardio-preventive health examination programme	136	Observational, within-subjects	1		Time point in study (baseline vs at 12 months)	(+) There is reactivity to measurement of LPA and SB but not MVPA.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
												The effect was moderated by time point in the study: measurement reactivity was stronger at second time point compared to the first.
Vanhelst et al. (2017)	France	PA (activity counts, LPA time, MPA time, MVPA time, VPA time), SB (sedentary time)	accelerometer (ActiGraph GT3X)	4	adolescents (10-18 years)		78	Experimental, between-subjects	2	unaware of use of device: device concealed as body posture monitor; aware of use of device	Schooldays vs school-free days	(-) There is no reactivity to measurement of any of the included indicators for PA. There were no differences between schooldays and school-free days.

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
Yang et al. (2015)	US	Alcohol consumption (alcoholic drinks per day)	smartphone app (emocha, Android)	At least 24	adults	African American men who have sex with men	15	Observational, within-subjects	1			(+) There was a trends towards reactivity to measurement of alcohol consumption that flattened out after 25 days.
Zhu and Haegele (2019)	US	PA (MVPA)	accelerometer (ActiGraph GT3X)	4	children, adolescents (6-17 years), adults (parents)	children with visual impairments, their siblings and parents	66	Observational, within-subjects	1		Visual disability (yes/ no), age (parents vs children)	(+) There is reactivity to measurement of MVPA in children and adults without visual disabilities. The direction of the effect differed: children without disabilities and their

Study	Country	Target behaviours	Assessment tool(s)	Study duration in days	Sample			Study design	Conditions		Moderators of measurement reactivity effect included in analysis	Conclusion [a]
					Age group(s)	Specific characteristics	N		Number	Description (if > 1)		
												parents showed an initial elevation of MVPA, while children with visual disabilities showed an initial decline.
Zhu et al. (2020)	China	PA (LPA, MVPA)	accelerometer (ActiGraph GT3X)	7	Adolescents (mean age 13.9)	with moderate to severe intellectual disability	175	Observational, within-subjects	1		Severity of intellectual disability	(+) There is reactivity to measurement of LPA and MVPA for both disability groups.

Notes. [a] (+) measurement reactivity occurred, (-) measurement reactivity did not occur; [b] Partial overlap in the reported data. Abbreviations: EMA – Ecological Momentary Assessment, LPA – light physical activity, MEMS - Medication Event Monitoring System, MPA – moderate physical activity, MVPA – moderate to vigorous physical activity, PA – physical activity, SB – sedentary behaviour, VPA – vigorous physical activity

Study design

A range of study designs was used to investigate reactivity to digital measurement of health behaviour. The majority of studies (68%, n = 21) (Baumann et al., 2018; Cook et al., 2012; Craig et al., 2010; Davis & Loprinzi, 2016 Dössegger et al., 2014; Haegele et al., 2020; Hilgenkamp et al., 2012; Klenk et al., 2019; Labhart et al., 2020; Ling & King, 2015; Ling et al., 2011; Motl et al., 2012; Motl & Dlugonski, 2011; Poulton et al., 2019; Sutton et al., 2014; Tinlin et al., 2018; Ullrich et al., 2021; Yang et al., 2015; Zhu et al., 2020; Zhu & Haegele, 2019) used observational within-subjects designs to test whether behaviour changed across the study period. Typically, a statistically significant change in behaviour was interpreted as measurement reactivity, e.g., a decline in steps or an increase in sedentary time, reflecting an initial elevation (or reduction) due to reactivity and a gradual return to pre-assessment levels. Studies either presented tests of linear effects across the study period, treating time as a continuous variable (e.g., Labhart et al., 2020; Poulton et al., 2019; Ullrich et al., 2021), or compared behaviour on the first study day to behaviour on a varying number of subsequent days (e.g., Davis & Loprinzi, 2016; Haegele et al., 2020).

The remaining ten studies compared measurement reactivity in at least two different conditions, either between (n = 4; Ho et al., 2013; McCarthy et al., 2015; Scott et al., 2014; Vanhelst et al., 2017) or within (n = 6; Clemes et al., 2007; Clemes & Parker, 2009; Clemes & Deans, 2012; Foley et al., 2011; Foote et al., 2017; Prewitt et al., 2013) participants. Typically, different types of devices were compared. Studies typically hypothesised that unobtrusive recording would lead to little or no measurement reactivity, while being aware of the device's purpose or even being able to see the recorded data would increase healthy behaviours (e.g., physical activity) or decrease unhealthy behaviours (e.g., sedentary behaviour). Five experimental studies additionally tested whether behaviour changed across the study period, e.g., between different days or weeks of the study (Clemes & Deans, 2012; Foley et al., 2011; Foote et al., 2017; Ho et al., 2013; Scott et al., 2014).

Types of devices compared in experimental studies

The ten experimental studies used a range of (manipulated) devices to study reactivity. The majority of studies used a device for which the actual use was concealed (Foote et al., 2017); e.g., Clemes and Parker (2009) concealed the pedometer as a body posture monitor (see also Clemes et al., 2007; Clemes & Deans, 2012; Vanhelst et al., 2017). In three of these studies, the use of the device was concealed in the first part of the study before being revealed for the second part of the study (Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009). In four studies, behaviour was compared between wearing a sealed (i.e., no data visible on the device) and an unsealed tracking device (Clemes & Parker, 2009; Foote et al., 2017; Prewitt et al., 2013; Scott et al., 2014). In three studies, participants were also asked to copy the feedback from the device into a diary (Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009). Two other studies provided participants with a concealed (Foley et al., 2011) or sealed (Ho et al., 2013) device for monitoring and provided some participants with a second device for which the purpose of recording physical activity was known and the data visible. Finally, one study compared low vs high frequency sampling conditions (McCarthy et al., 2015).

Measurement reactivity in digital assessment of health behaviour

Physical activity

For seven experimental studies (Clemes et al., 2007; Clemes & Parker, 2009; Foley et al., 2011; Ho et al., 2013; Prewitt et al., 2013; Scott et al., 2014; Vanhelst et al., 2017), effect sizes for the comparison of measurement reactivity manipulations could be obtained. First, a random-effects (RE) meta-analysis was conducted combining all studies independent of the manipulation or physical activity indicator (see Figure 2), yielding a small and statistically significant pooled effect size for Cohen's $d = 0.27$, 95% CI [0.16; 0.39] (test for overall effect: $Z = 4.58$, $p < .001$). There was substantial heterogeneity as indicated by I2 of 78% (Tau2 =

0.04, H2 = 4.60, df = 17, p < .001; The Cochrane Collaboration, 2011a). Due to this heterogeneity, results of the narrative synthesis are reported separately per physical activity indicator. Another separate meta-analysis was conducted for steps; it was not possible to conduct further meta-analyses due to the small number of experimental comparisons. Asymmetry was examined using a funnel plot (see Figure 3) and Egger's test (Egger et al., 1997; Sterne & Egger, 2005), which was not significant (p = .976), thus not providing evidence for publication bias. The funnel plot, adjusted using the trim and fill method (Duval & Tweedie, 2000), is presented in Figures 5, Appendix D. The effect size estimate remained unchanged after trim and fill.

Steps/walking. The majority of studies targeting physical activity recorded participants' step counts or stepping/ walking time using pedometers and accelerometers. Ten of a total of 15 studies found evidence for reactivity to the measurement. Evidence stems from both observational (Hilgenkamp et al., 2012; Ling et al., 2011; Motl & Dlugonski, 2011; Motl et al., 2012; Tinlin et al., 2018) and experimental designs (Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009; Scott et al., 2014). Four observational (Craig et al., 2010; Haegele et al., 2020; Klenk et al., 2019; Ling & King, 2015) and one experimental study (Prewitt et al., 2013), on the other hand, did not find evidence for measurement reactivity. One observational study additionally reported evidence for reactivity to the measurement of stepping time, but not for standing time (Tinlin et al., 2018). A random effects meta-analysis of the four experimental studies (Clemes et al., 2007; Clemes & Parker, 2009; Prewitt et al., 2013; Scott et al., 2014) on steps (see Figure 3) yielded a statistically significant small to medium pooled effect size for Cohen's d = 0.30, 95% CI [0.12; 0.49] (test for overall effect: Z = 3.20, p = 0.001). Again, heterogeneity was substantial (I2 = 86%, Tau2 = 0.07, H2 = 7.27, df = 9, p < .001; The Cochrane Collaboration, 2011a). Asymmetry was investigated using a funnel plot (see Figure 3) and Egger's test (Egger et al., 1997; Sterne &

Egger, 2005), which was not significant (p = .475), thus not providing evidence for publication bias. The funnel plot adjusted using the trim and fill method (Duval & Tweedie, 2000) is presented in Figure 6, Appendix D. Applying trim and fill increased the pooled effect size to Cohen's d = 0.39, 95% CI [0.19; 0.59], Z = 3.91, p = .001.

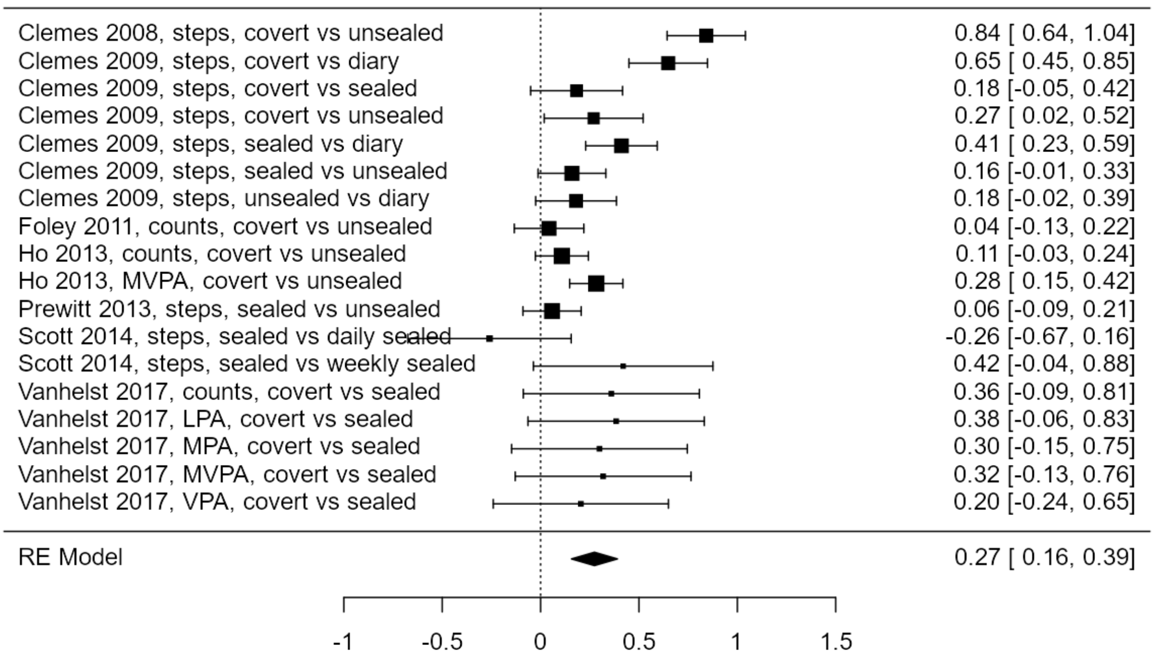


Figure 2. Forest plot of experimental studies with all physical activity outcomes. Effects on the right side of the dashed line indicate that manipulations that were hypothesised to increase measurement reactivity did indeed increase physical activity. Effects on the left side of the dashed line indicate the opposite effect. RE = random effects.

MVPA. A total of nine studies assessed moderate and/or vigorous physical activity using accelerometers. Two of these studies found a significant decrease in MVPA across the study period using an observational design, which is in line with the hypothesised effect (Zhu et al., 2020; Zhu & Haegele, 2019). The remaining seven studies, including two experimental studies (Ho et al., 2013; Vanhelst et al., 2017), did not report measurement reactivity in their data (Baumann et al., 2018; Davis & Loprinzi, 2016; Haegele et al., 2020; Ho et al., 2013; Tinlin et al., 2018; Ullrich et al., 2021; Vanhelst et al., 2017).

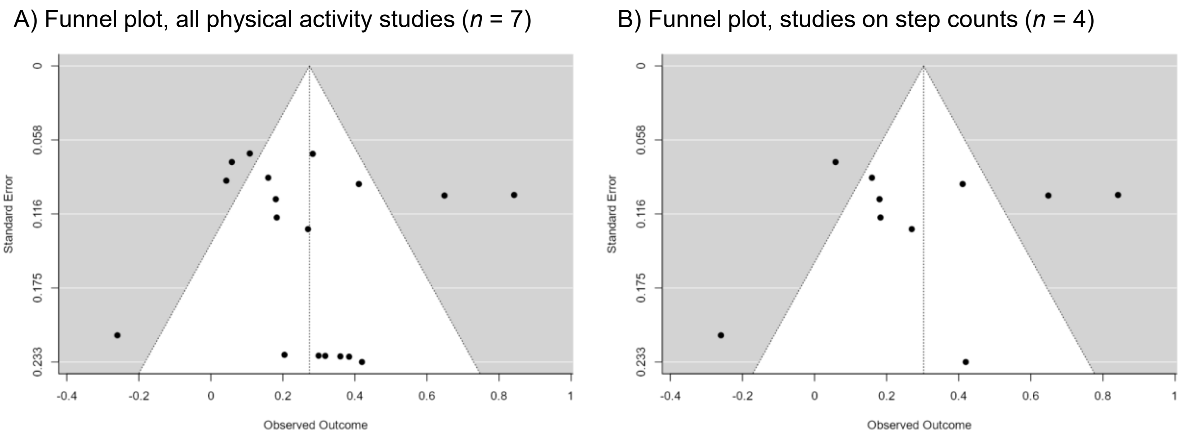


Figure 3. Funnel plots for all experimental studies investigating reactivity to measuring physical activity (panel A) and for all studies investigating reactivity to measuring step counts (panel B).

LPA. Six studies assessed light physical activity using accelerometers. Three observational studies concluded that measurement reactivity occurred (Baumann et al., 2018; Ullrich et al., 2021; Zhu et al., 2020), while another observational study and one experimental study concluded that there was no reactivity to the assessment (Tinlin et al., 2018; Vanhelst et al., 2017). However, the non-significant effect in the experimental study was small to medium (Cohen's d of 0.38). One observational study reported changes in the data across the study period that were in line with the hypothesized measurement reactivity effect. The effect did not reach statistical significance but exceeded five percent change (Haegele et al., 2020). (Figure 4).

Counts. Seven studies assessed physical activity using activity counts provided by accelerometers. Two observational studies reported a decline in activity counts over time which was interpreted as measurement reactivity (Dössegger et al., 2014; Motl & Dlugonski, 2011). Three experimental studies also reported changes in activity counts depending on the condition; participants recorded higher activity counts if the counts were visible to them (Foley et al., 2011; Foote et al., 2017; Ho et al., 2013). The reported effects, however, were very small: Cohen's d ranged from 0.04 to 0.11. Foley et al. (2011) reported an initial elevation of activity counts in the first thirty minutes of wearing the unsealed device, which attenuated afterwards. On the other hand, one observational (Davis & Loprinzi, 2016) and one experimental study (Cohen's d of 0.38; Vanhelst et al., 2017) concluded that measurement reactivity did not occur.

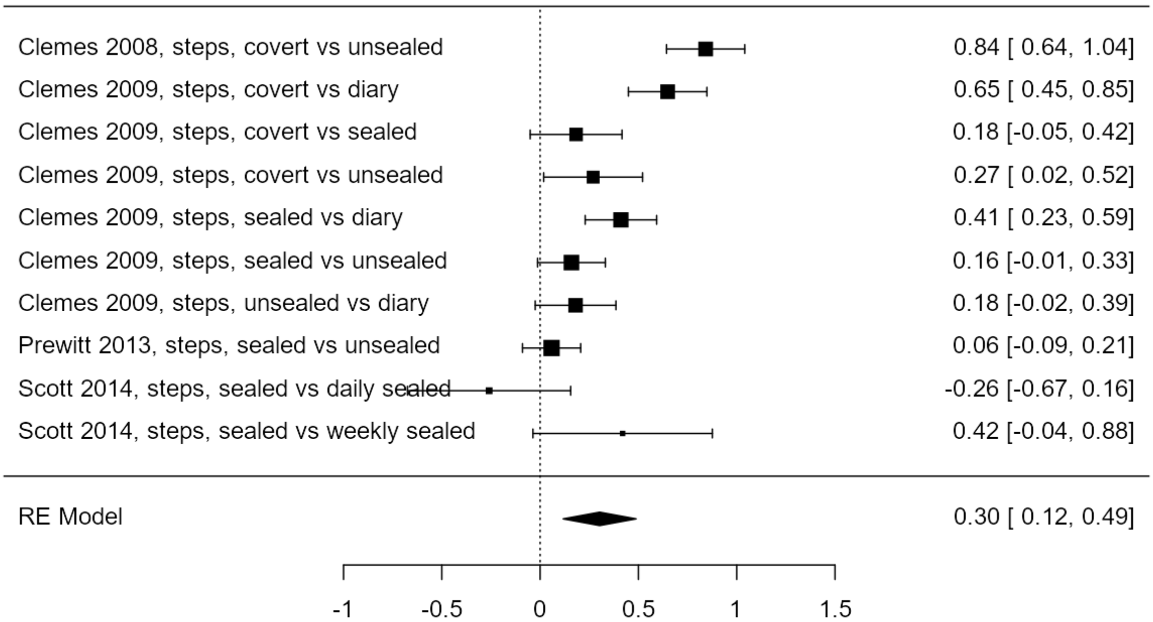


Figure 4. Forest plot of experimental studies focusing on steps. Effects on the right side of the dashed line indicate that manipulations that were hypothesised to increase measurement reactivity did indeed increase the number of recorded steps. Effects on the left side indicate the opposite direction. RE = random effects.

Sedentary behaviour

Four studies assessed sedentary time using accelerometers, which again reported mixed findings regarding the presence of measurement reactivity. Two observational studies reported a decline in sitting time across the study period, which was taken as an indicator for measurement reactivity (Baumann et al., 2018; Ullrich et al., 2021). Two other studies, one of which was experimental, concluded that measurement reactivity did not occur (Tinlin et al., 2018; Vanhelst et al., 2017). Although not statistically significant, the group difference in the experimental study indicated a small to medium effect of Cohen's d = 0.42.

Alcohol consumption

Two out of three observational studies did not provide evidence of reactivity to digital measurement of the number of alcoholic drinks consumed using smartphone apps (Labhart et al., 2020; Poulton et al., 2019). Yang et al. (2015), on the other hand, reported a slight decrease in the number of drinks which levelled off at 25 days.

Consumption of non-alcoholic drinks

One observational study that focused on alcohol consumption also reported the number of nonalcoholic drinks consumed. This study did not find evidence for reactivity to the measurement (Labhart et al., 2020).

Medication adherence

One observational study and observational data from a larger RCT on medication adherence both reported some measurement reactivity to using electronic monitoring systems; the effect only reached statistical significance in Cook et al. (2012). They specifically point towards a drop in adherence between five and six weeks of recording. Sutton et al. (2014) also report a small albeit nonsignificant decline in adherence across the study period.

Smoking

One study investigated reactivity to digital assessment of smoking cessation and tobacco abstinence using EMA devices (McCarthy et al., 2015). In this study, the prompting frequency was manipulated; some participants received one prompt per day to report smoking while other participants were prompted six times per day. The prompting frequency did not impact cessation or abstinence. The authors thus concluded that prompting frequency does not impact reactivity to the measurement.

Moderators

Seventeen studies included at least one moderator. There was substantial heterogeneity in the potential moderators of measurement reactivity explored in the included studies. No significant effects were found for whether activities in vs out of school were recorded (Foote et al., 2017; Vanhelst et al., 2017), the participants' Body-Mass Index, employment status (Clemes & Deans, 2012), and the number of days on which a report was completed (Labhart et al., 2020). Significant effects were found for the time point within a longitudinal study, with more reactivity occurring at the second compared to the first time point (Ullrich et al., 2021); the season in which the study was conducted with steeper increases in sedentary time and steeper decreases in LPA when data was collected in summer vs winter or spring; whether recording started on a weekday or at the weekend, with stronger declines in MVPA if measurement started on a weekday (Baumann et al., 2018); the time of day when the recording took place, with measurement reactivity occurring earlier but not later in the day (Foley et al., 2011), visual disabilities, with children with visual disabilities showing an initial decline in MVPA while children and adults without visual disabilities showed an initial increase (Zhu & Haegele, 2019), the tendency to ruminate, with more pronounced reactivity in participants with a stronger tendency to ruminate (Ling et al., 2011); and hazardous drinking behaviour, with the effect only being present in participants who engaged in hazardous drinking behaviour, but not in participants who did not engage in hazardous drinking behaviour (Poulton et al., 2019).

Indicators for the awareness of being measured

Six studies included moderators that reflected participants' awareness of being measured and their understanding of what participating in a study entails. Young children, for instance, may exhibit less social desirability bias due to lack of awareness. Accordingly, four studies included participants of different age groups ranging from children to adults (Davis & Loprinzi, 2016; Dössegger et al., 2014; Prewitt et al., 2013; Zhu & Haegele, 2019). Davis and Loprinzi (2016) reported a stronger measurement reactivity effect in adults vs children, while Dössegger et al. (2014) reported measurement reactivity to occur in children aged 7 years and older, but not in children between the age of 3 and 6. Zhu and Haegele (2019), on the other hand, did not report systematic differences between children and adults. Similarly, Prewitt et al. (2013) did not report differences between school children in grades 4, 5 or 6. Finally, Hilgenkamp et al. (2012) compared effects in adults under and over the age of 65, reporting no significant differences.

Another moderator included in two studies was intellectual disability (Hilgenkamp et al., 2012; Zhu et al., 2020), since it was hypothesised that people with intellectual disabilities may lack the understanding of the implications of measurement as part of a study. This hypothesis, however, was not confirmed: the strength of measurement reactivity did not differ between participants with low and high levels of intellectual disabilities.

In a similar vein, Prewitt et al. (2013) tested for the moderating role of knowledge about pedometers. Again, no significant effect was found.

Gender

Six studies investigated gender differences in measurement reactivity. Five studies did not find significant differences between male and female participants (Clemes & Deans, 2012; Clemes & Parker, 2009; Davis & Loprinzi, 2016; Hilgenkamp et al., 2012; Prewitt et al., 2013). Ho et al. (2013), on the other hand, found reactivity to measurement of activity counts in girls, but not in boys.

Risk of bias assessment

Risk of bias was assessed using three different tools, depending on the study design. Results are summarised in Tables 3–5 in Appendix D. For four studies using a randomised between-subjects design (Ho et al., 2013; McCarthy et al., 2015; Scott et al., 2014; Vanhelst et al., 2017), the Cochrane Risk of Bias 2.0 tool was used (Sterne et al., 2019). All four studies were subject to significant risk of bias (see Table 2 for details), with two studies receiving the overall rating of some concerns (Scott et al., 2014; Vanhelst et al., 2017) and two receiving the overall rating of high risk of bias (Ho et al., 2013; McCarthy et al., 2015).

The high risk of bias arose from a lack of blinding, which was impossible due to the used assessment tool. Furthermore, since none of the four studies reported a prespecified analysis plan, some risk of bias arose from the reported result.

Risk of bias of six studies using a within-subjects design (Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009; Foley et al., 2011; Foote et al., 2017; Prewitt et al., 2013) was evaluated using the checklist for cross-over studies by Ding et al. (2015). This checklist does not provide a summary evaluation. For all but one study, high risk of bias arose from non-randomised order of treatments (Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009; Foote et al., 2017; Prewitt et al., 2013); randomising the treatment order in these studies was not possible since they relied on participants being blinded to the use of the measurement device. Furthermore, no study specifically addressed potential carry-over effects of the treatments, resulting in unclear risk of bias. Similarly, for all six studies potential risk of bias arose from a lack of information on blinding participants and the research team. Finally, other potential sources of bias could not be assessed for all studies due to a lack of information.

Twenty-one observational studies were appraised using the JBI Checklist for Analytical Cross Sectional Studies (Moola et al., 2020), which again does not provide a summary evaluation. Some issues arose regarding defining inclusion criteria in six studies (Craig et al., 2010; Dössegger et al., 2014; Ling et al., 2011; Ling & King, 2015; Ullrich et al., 2021; Zhu et al., 2020) as well as regarding identifying and addressing confounding factors in nine (Klenk et al., 2019; Labhart et al., 2020; Ling et al., 2011; Ling & King, 2015; Motl et al., 2012; Motl & Dlugonski, 2011; Tinlin et al., 2018; Yang et al., 2015)/ eight studies (Klenk et al., 2019; Ling et al., 2011; Ling & King, 2015; Motl et al., 2012; Motl & Dlugonski, 2011; Tinlin et al., 2018; Yang et al., 2015), respectively.

2.5. Discussion

Summary of main findings

Digital in-the-moment assessment of health behaviour has become increasingly popular in recent years, mainly because it is believed to suffer less from typical shortcomings of self-report research such as recall bias (Thompson & Subar, 2017). The present systematic review aimed to investigate the validity of health behaviour data collected in-the-moment with digital devices by synthesising the literature on reactivity to the measurement. The majority of identified studies focused on different aspects of physical activity. Overall, evidence for measurement reactivity was mixed. Effect sizes derived from experimental studies showed large heterogeneity and ranged from very small to large. However, the overall effect identified in the meta-analysis was small to medium (c.f. Cohen, 1992), indicating that reactivity to the measurement of physical activity exists to some extent. The results may be useful when modelling research participation effects to quantitatively account for them in the data analysis without needing to formally assess measurement reactivity in individual studies (see Bendtsen & McCambridge, 2021, for recommendations).

The results of this systematic review indicate that measurement reactivity is not limited to selfreports in questionnaires but also extends to digitally assessed behavioural data. Indeed, confidence intervals of the meta-analysis suggest that reactivity for digitally assessed physical activity might be at least as strong, if not stronger, than for questionnaire-based assessments: previous meta-analyses produced pooled effect sizes (standardised mean difference, Hedge’s g) of 0.19–0.21 (Miles et al., 2020; Wilding et al., 2016). This observation is in line with a previous meta-analysis that also reported stronger effects for objective compared to self-report measures, although this difference was not statistically significant (Wood et al., 2016). These findings challenge the assumption that objective measures of behaviour may lead to more ecologically valid conclusions than self-report measures; however, additional benefits of digital in-the-moment assessments such as reduced recall bias and the opportunity to study behaviour repeatedly and in daily life (Trull & Ebner-Priemer, 2013) remain.

Several studies included in this review suggest that measurement reactivity might be short-lived: Reactive effects are most likely to occur in the first hours to days of assessment and attenuate afterwards. It may thus be advised to exclude the data from the first days of use. For instance, Clemes and Deans (2012) and Foote et al. (2017) suggest to exclude data collected on physical activity from the first week of recording. Recommendations, however, may differ between behaviours; one study on medication adherence suggests to exclude data from the first six weeks of use (Cook et al., 2012), which may not always be feasible.

This systematic review highlights a lack of research on reactivity to the measurement of behaviours other than physical activity. Consumption behaviours such as medication adherence, alcohol consumption, smoking, and eating behaviour were rarely studied, which may be the case because they are more difficult to assess digitally than physical activity, for which passive tracking is well established. Most notably, although digital assessment of dietary intake including objective indicators such as photos is common in behavioural research (König et al., 2021), measurement reactivity has not yet been studied in this domain: Apart from one study that investigated the consumption of both alcoholic and non-alcoholic drinks during nights out, no studies on dietary intake were identified. Future research needs to address this gap to provide important information on the validity of digital assessment of dietary intake as well as other consumption behaviours.

Explanations for measurement reactivity

Whether measurement reactivity occurs might depend on the indicator of the target behaviour, which varied substantially especially in the studies investigating physical activity. For instance, a large proportion of studies that investigated steps or activity counts as indicators for physical activity, which have been shown to correlate highly (Sartini et al., 2015), reported reactive effects. Studies investigating moderate to vigorous physical activity, on the other hand, rarely reported reactive effects. Steps and activity counts may be more easily modifiable than moderate to vigorous physical activity since they require less effort and are easier to integrate in existing daily life activities (Cheval & Boisgontier, 2021). For instance, it may be easier to increase the number of steps by getting off the bus one stop early and walking the rest of the way than to schedule a formal exercise session at the gym on an already busy day. Only a small number of studies included in this review allowed for a direct comparison of different physical activity indicators by assessing several indicators in the same sample. The only study that included both steps and activity counts found reactive effects for both indicators (Motl & Dlugonski, 2011). Moreover, Ho et al. (2013) and Tinlin et al. (2018) simultaneously assessed steps or activity counts and moderate to vigorous physical activity and reported reactivity to the measurement of the former, but not the latter indicator. Although they did not formally test this hypothesis, they support the notion that measurement reactivity is more likely to occur when the behaviour is easier to modify. On the other hand, Vanhelst et al. (2017) reported reactive effects for neither indicator. Future research on measurement reactivity thus should explicitly take several indicators for the same behaviour into account that differ in the required effort to explicitly test this hypothesis.

This systematic review also investigated a large number of potential moderators of measurement reactivity. Based on the hypothesis that measurement reactivity is caused by forming beliefs about the study team's expectations that translate into behaviour via social desirability (McCambridge et al., 2014), several studies tested potential moderating effects of awareness of the measurement's purpose. Some studies experimentally manipulated awareness by concealing the purpose of the device (Clemes et al., 2007; Clemes & Deans, 2012; Clemes & Parker, 2009; Foley et al., 2011; Ho et al., 2013; Vanhelst et al., 2017).

Results from experimental studies were inconclusive, which could be explained by heterogeneity in the comparators (e.g., sealed or unsealed devices). Moreover, most studies used samples of less than 100 participants, which is too small to reliably detect the small to medium effects that the majority of studies reported. Other studies specifically recruited participants who were expected to be unaware of the implications of taking part in a study, such as children or people with intellectual disabilities (Davis & Loprinzi, 2016; Dössegger et al., 2014; Hilgenkamp et al., 2012; Prewitt et al., 2013; Zhu & Haegele, 2019; Zhu et al., 2020). However, it is unclear whether those subsamples were in fact aware of the implications of measurement in a research project since awareness was rarely assessed. A notable exception is Prewitt et al. (2013), however, they investigated a narrow age range with school children in the 4th, 5th and 6th grade. The results of this review regarding awareness as a potential underlying mechanism thus do not necessarily challenge this hypothesis, but rather underline the need for studying larger samples and choosing comparators carefully in future research.

Similarly, familiarity with tracking health behaviours might impact whether measurement reactivity occurs in the context of research. Tracking physical activity using wearables or smartphone apps has become popular in recent years (König et al., 2018; Pew Research Center, 2020), so one might speculate that regular trackers may not experience increased awareness at the beginning of a study and may thus be less prone to measurement reactivity. This assumption, however, is yet to be tested in empirical research.

Some of the included studies suggest yet another explanation for measurement reactivity. For instance, several studies compared sealed vs unsealed devices. In both conditions, participants were aware of the measurement, but only when using an unsealed device, they also had access to the data that was collected. Having access to the data may serve as a reminder of study participation; however, it may also induce other cognitive processes such as reflection on the current behaviour (Barta et al., 2012). This may lead to detecting a discrepancy between the current and the desired behaviour, which in turn may trigger self-regulatory processes to adjust the behaviour (Carver & Scheier, 1982; Kanfer & Gaelick-Buys, 1991). Accordingly, when planning studies, researchers might need to avoid assessment tools such as accelerometers or fitness trackers that display the recorded data, or may need to take additional precautions so that the data are not visible, to reduce reactive effects (French et al., 2021b). This may be especially important when using low-cost trackers that typically have displays (Degroote et al., 2020b). Although not all participants might necessarily have the intention to change behaviour when they enrol in the study, participants of health-related studies often are more health-conscious than the population average due to self-selection bias (also referred to as volunteer bias, see e.g., Haynes & Robinson, 2019 and Nuzzo, 2021; for discussions) and thus may also be more likely to have the intention to change their behaviour in line with recommendations. Thus, researchers might need to assess intention to change behaviour and introduce this variable as a covariate in the analysis to investigate potential intra-individual differences in measurement reactivity that may arise from differences in intention.

In a similar vein, Poulton et al. (2019) investigated whether reactivity was stronger in participants who showed unhealthy alcohol consumption patterns; accordingly it could be hypothesised that measurement reactivity might be more pronounced in individuals wh behave in a comparably unhealthy way and thus may see greater need for change (see also Barta et al., 2012, for a summary). Indeed, only participants with hazardous drinking behaviour showed reactivity to measuring alcohol intake. Future research is needed to confirm these findings and to test their generalizability to other behaviours.

Limitations

It is important to address several shortcomings of this review and the included studies. Although a few studies with several hundred participants were included that were powered to detect small effects which would have been expected based on previous reviews on questionnaire-based research (Miles et al., 2020; Wood et al., 2016), half of the included studies had less than 100 participants. Accordingly, uncertainty of the reported effects is large, as indicated by 95% confidence intervals of the pooled effects ranging from very small to medium. Especially studies with small samples might have missed small effects. Still, even small effects may have important implications for behavioural research (Götz et al., 2022), in which the mean effect size is small to medium (Cumming & Calin-Jageman, 2017), as well as clinical trials (French et al., 2021b). Accordingly, researchers planning further studies on reactivity to digital measurement should account for potential small effects by recruiting sufficiently large samples. Moreover, future research needs to address shortcomings in study quality as outlined by quality appraisal tools; common issues included a lack of or missing information on blinding. Furthermore, no study used a pre-specified analysis plan; this can be solved through pre-registration.

Moreover, the studies included in the meta-analysis were heterogeneous in terms of a range of indicators for physical activity as well as study designs and compared devices. It was thus not possible to conduct meta-analyses separately for each physical activity indicator other than steps, and also the different manipulations could not be separated. More research is needed on systematically comparing different manipulations that may induce measuremen reactivity to allow for the effects to be synthesised separately.

Two thirds of the studies included in this review were observational and they operationalised measurement reactivity as a particular pattern of change in behaviour across time. Observational study designs provide important insights into the temporal dynamics of measurement reactivity; however, changes in behaviour across time may also be induced by other influences such as participant fatigue and subsequent gaps in recording (Ziesemer et al., 2020). Experimental designs are more robust, since they allow for the direct and controlled comparison of different conditions, but they may not always be feasible. For instance, it is possible to conceal the true purpose of certain devices to record physical activity; however, unobtrusive recording is typically not feasible or ethical for consumption behaviours (e.g., when using sensors that are applied to the participant's body, see Bell et al., 2020, for an overview). Still, future research may provide important insights into key components of the measurement tool that amplify reactivity, e.g., by comparing measurement tools that differ in burden or complexity (French et al., 2021b).

Lastly, due to the difficulties of measuring indicators of consumption behaviours (e.g., alcohol consumption, smoking, eating/ drinking) fully objectively, this review also included studies that investigated reactivity to digital recording of behaviour in-the-moment or close to consumption. This caveat is important to keep in mind when interpreting the findings. However, through deviating from the pre-registered eligibility criteria for this review, a more complete picture of measurement reactivity research regarding digital assessments was obtained.

2.6. Conclusions

In summary, most research on reactivity to digital in-the-moment measurement of health behaviour focuses on physical activity. Measurement reactivity effects are generally small, but meaningful. The results extend a recently published list of study features that may indicate risk of bias due to measurement reactivity (French et al., 2021b). For instance, measurement reactivity may be amplified when studying behaviours that require comparably little effort to change, such as the number of steps. Researchers using digital in-the-moment assessment tools might want to focus on moderate to vigorous physical activity, use assessment tools that do not provide participants access to the data, and use a run-in period of several days to a week to minimise reactive effects. More research is needed especially on potential reactivity to the assessment of consumption behaviours to be able to provide further guidance for researchers. Future studies should use experimental designs which enable different assessment methods to be compared and thus to identify the methods that induce least measurement reactivity. In this way, the validity of health behaviour research and thus the effectiveness of health promotion programs can be improved.

3. The same, only different: Smartphone-based dietary Ecological Momentary Assessment tools vary in complexity, usability, and active information processing.

Anila Allmeta [1], Stephen Sutton [2], Laura M. König [1,3]

[1] Faculty of Life Sciences: Food, Nutrition and Health, University of Bayreuth, Germany
[2] Department of Public Health and Primary Care, University of Cambridge, United Kingdom
[3] Faculty of Psychology, University of Vienna, Austria

This chapter has been published in:

Allmeta, A., Sutton, S., & König, L. M. (2026). The same, only different: Smartphone-based dietary Ecological Momentary Assessment tools vary in complexity, usability and active information processing. *British Journal of Health Psychology*, *31*(1), e70057. https://doi.org/10.1111/bjhp.70057

<u>Contribution:</u>

AA: conceptualization, data curation, writing – original draft, formal analysis, investigation, Methodology, project administration; SS: conceptualization, methodology, writing-review & editing; LK: conceptualization, methodology, writing-review & editing.

3.1. Abstract

Objectives: Ecological Momentary Assessment (EMA) is popular for assessing dietary intake in real life and real time. Available tools differ substantially in the type and number of implemented features, including features to assess what and how much was consumed. These features require qualitatively different input that might have a differential impact on the participants' cognitions and behaviours while taking part in the study. This study aimed to test whether more complex dietary assessment tools, indicated by the type and number of assessment features, induce more active information processing (AIP).

Design: Preregistered online between-subjects experimental study.

Methods: A total of 373 participants (65.4% female; mean age 30.4 years) were randomly allocated to view one of eight EMA protocol mock-ups, each describing a food tracking process verbally and using screenshots. Afterwards, they rated the protocol in terms of its complexity (manipulation check), AIP and its potential impact on eating-related cognitions, intentions and eating behaviour change.

Results: The eight EMA protocols differed in perceived complexity, that is protocols with more tracking features were perceived as more complex compared to those with fewer tracking features. EMA protocols that were perceived to be more complex were also perceived to induce more AIP. However, there were no differences in perceived impact on eating-related cognitions, intentions and behaviour.

Conclusions: Differences in complexity and usability may influence compliance and study results. Researchers thus need to carefully select the appropriate EMA protocol for their study to balance the need for collected information with the need for high compliance.

3.2. Introduction

Smartphone-based dietary assessment via Ecological Momentary Assessment (EMA) (Shiffman et al., 2008) is becoming increasingly popular in health psychology and related fields to assess a range of health behaviours (Ottenstein & Werner, 2022; Perski et al., 2022 ; Wrzus & Neubauer, 2023), including dietary behaviours (König, van Emmenis et al., 2022). It enables food intake in daily life to be assessed without recall bias (Sharp & Allman-Farinelli, 2014), in relation to various determinants (Perski et al., 2022) and across time, which may lead to more accurate measurement and more robust findings, and subsequently may advance psychological theories (Scholz, 2019).

A multitude of smartphone-based dietary assessment tools have been used in research that differ substantially in the type and number of implemented features (König, van Emmenis, et al., 2022). For instance, some tools require participants to take photos (e.g., König and Renner, 2018), some ask participants to search for individual food items in a database (Ambrosini et al., 2018) or through classification systems (Biel et al., 2018; Elliston et al., 2017) and indicate portion sizes (Béjar et al., 2018; Jung et al., 2020), and others ask for a free-text description of the food intake (König & Renner, 2018, 2019).

These features require qualitatively different input. It could be hypothesized that they also exert differential impact on participants' cognitions and behaviours while taking part in the study. Indeed, prior research both in the context of surveys and intensive longitudinal assessments suggests that study participants' behaviours, cognitions or emotions may change due to study participation, a phenomenon known as measurement reactivity (French & Sutton, 2010). Effects are often small, but may still be meaningful, given that effects of behavioral interventions also typically report small to medium effect sizes (König et al., 2022; Miles et al., 2020; Maher et al., 2024). Studies specifically testing reactivity to the measurement of dietary intake are sparse. Research in survey studies so far produced highly heterogeneous findings, and according to a recent systematic review, the issue has not been addressed at all in intensive longitudinal studies on diet. The literature furthermore speculates about the potential role of involving participants actively in the recording process. For example, one study found stronger reactivity effects if participants were explicitly asked to copy the step count from their step counter into a paper diary compared to when they were not (Clemes & Parker, 2009).

When recording dietary intake, the different tools also require different interactions with the recording tools, which may induce different levels of active processing of information such as sensory input while recording. In other contexts, more active processing of information is related to better understanding of information and performance (e.g., Natter and Berry, 2005). It could thus be assumed that more active information processing in dietary assessment leads to more pronounced changes in eating behaviour (see also Barta et al., 2012). For example, selecting individual foods from a database and adding portion sizes requires study participants to provide detailed information on their meal, which requires active reflection. Taking a picture of the meal, however, can be completed in seconds and might require less active reflection. Thus, depending on which assessment tool is used to record food intake, active information processing and reflection on the food choices might vary substantially. In the present study, active information processing is defined as the active reflection on the information that the participants have to provide while recording their food intake using different recording features (i.e., photo-based, food database, detailed description). If these differences indeed translate into behaviour, researchers would need to be careful in comparing data from studies that used dietary assessment tools with different features.

The present study provides a first test of these assumptions. Specifically, the study aimed to test whether more complex dietary assessment tools, where complexity is indicated by the type and number of assessment features, induce more active information processing. We hypothesized that assessment tools that, for instance, require participants to describe individual components by searching in a food database and input serving sizes, are perceived (i.e., rated by the participants) to be more complex (H1), to induce more active information processing (H2) and to subsequently produce greater changes in eating-related cognitions, intentions, and behaviour (H3) (c.f., French & Sutton, 2010) than less complex tools (e.g., taking a photo only). Furthermore, the study aimed to explore the usability of the different dietary assessment tools (exploratory research question; E1), which is more often reported for assessment or intervention apps (König et al., 2021) and linked to complexity of tools; and to test whether previous experience with nutrition apps (E2) and the level of affinity to technology (E3), i.e. an individual's tendency to actively engage with technology (Franke et al., 2019), influenced participants' perceptions on active information processing, eating-related cognitions, intentions, and behaviours. Taking these individual characteristics into account is important since health behaviour tracking through EMA has become popular in recent years (König et al., 2018; Pew Research Center, 2020), and experience and affinity to technology may influence engagement with the tracking device (Seifert et al., 2017) and thus moderate the postulated effects.

3.3. Methods

This study was preregistered on the Open Science Framework (OSF; https://osf.io/bk27p), and all study materials and data are available from the OSF (https://osf.io/4htx7/). The study protocol was approved by the Ethics Committee of the University of Bayreuth.

Sample

A power analysis in GPower 3.1 (Faul et al., 2009) indicated a sample size of $N = 368$ for an alpha level of .05, and a small to medium effect size (Cohen's d=.4 or $\eta^2p = .06$) (Cohen, 1992; Lakens, 2013) using a one-way between-subjects ANOVA with eight groups for 80% power. Inclusion criteria were being at least 18 years old, having a good command of the English language, and having access to a smartphone/computer/tablet for the duration of the study.

A total of 373 participants completed the questionnaire; they were recruited through social media (i.e., official Instagram page of the faculty where the study was conducted; on the authors' professional and personal social media profiles on Instagram and Twitter/ X, special interest groups pages within the psychology community) and word of mouth in spring and summer 2021. The mean age of the participants was 30.4 years ($SD = 9.8$) with 65.4% women, 34.1% men, and 0.5% indicating another gender. The majority of the participants held a university degree (42.1% Bachelor's degree, 24.1% Master's degree). Regarding previous experience with nutrition apps, 56.3% stated to not have any experience, whereas 43.7% ($n = 163$) reported to have some experience (Table 1). The most frequently used nutrition apps were My Fitness Pal (39.2%), Yazio (11.9%), Fitbit (10.8%) and Lifesum (9.7%).

Table 1. Participants' prior experience with nutrition apps.

Stage	N	%
1: I have never thought about using an app for that.	101	27.1
2: I have thought about using an app for that, but so far, I did not do it	73	19.6
3: I have thought about using an app for that, but it is not necessary for me to do it	36	9.7
4: I am currently using an app for that and intend to continue to use it	38	10.2
5: I have used an app for that, but I do not use it anymore	125	33.5

Study design and procedure

An experimental online study with one between-subject factor (*EMA protocol*) was conducted. All participants provided informed consent prior to participating in the study. Participation was voluntary, anonymous, and participants did not receive any compensation or feedback. The online study was conducted using the survey platform Tivian Unipark and lasted for approximately 15 minutes. First, participants provided demographic information; then they were randomly allocated by the survey platform to view one of the eight EMA protocols, each describing a food tracking process reflecting a typical nutrition app in the form of mock-ups. The reporting process was both illustrated with a series of screenshots of the hypothetical app and a written description of each step. Afterwards, they rated the hypothetical assessment tool in terms of its complexity (manipulation check), active information processing and its potential impact on eating-related cognitions, intentions and eating behaviour change. Then, participants completed the short form of the User Experience Questionnaire (Schrepp et al., 2017). Finally, they provided information on their experience with nutrition apps and completed the Affinity to Technology (Franke et al., 2019) short scale (Wessel et al., 2019).

Materials and measures

EMA Protocols

Eight draft protocols were created in the form of mock-ups of typical nutritional apps that were identified based on the systematic review by König, van Emmenis et al. (2022). The mock-ups included screenshots of the assessment tool to show how the tool would look, as well as written descriptions of the recording process to explain how the tool should be used. The design of the mock-ups (i.e., colour scheme, icons, font etc.) was identical across the eight mock-ups for consistency. In the study, participants were asked to read the description, look at the screenshots and take their time to imagine that they would be using the nutrition app to record their food intake (breakfast, lunch, dinner, snacks, and drinks) several times per day, for several consecutive days.

All mock-ups included a first screenshot in which the user would hypothetically choose to record a meal (breakfast, lunch, dinner), a snack or a drink. The rest of the screenshots that followed, differed for the eight protocols; their components were based on the five core features of nutrition apps identified in König et al. (2022): photo-based assessment, food database, classification system, free-text description, assessment of serving or portion size. The combination of features in the different mock-ups in presented in Table 2. The eight mock-ups were assumed to differ in complexity. Initially, we conceptualised complexity via information load, for which the number of screenshots and number of words used to describe the recording process were used as pre-defined indicators (see Table 2.). Depending on the mock-up the participant was shown (thus also the tracking feature(s) combinations), they would hypothetically either take a photo of their meal, describe in detail the ingredients of the meal, search the meal in a food database, categorize the food into food categories (i.e., vegetables and fruits, starchy food etc..), or report the serving size (i.e., cups, portions or grams). Finally, in all protocols, the mock-up included the same last screenshot confirming the recording. Figure 1 illustrates EMA protocol 3, in which only the photo recording tracking feature was used, while Figure 2 illustrates EMA protocol 7, in which four features (photo recording, food database, detailed description, serving size) were combined. All screenshots are available from the project's OSF page.

Table 2. Description of the EMA protocol mock-ups used in the study, including the tracking features, number of screenshots and length of the description.

EMA protocol no.	Tracking features	Number of screenshots	Words
1	Detailed description	3	91
2	Classification system	3	95
3	Photo recording	3	112
4	Classification system + Serving size	4	134
5	Food Database + Serving size	4	152
6	Photo recording + Detailed description	4	159
7	Food Database + Detailed description + Serving size	7	318
8	Photo recording + Food Database + Detailed description + Serving size	8	383

Perceived complexity (manipulation check)

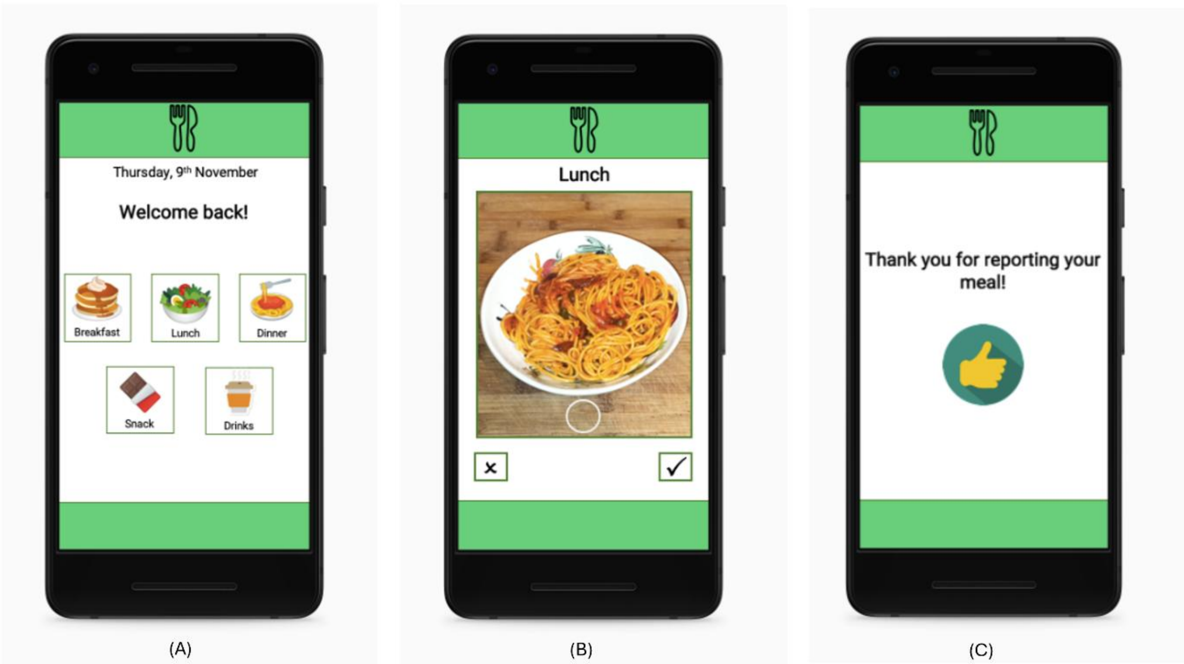
Participants were asked to rate the complexity of the EMA protocol they were shown on six items on a 5-point Likert scale (1-strongly disagree to 5-strongly agree). For instance, they rated whether recording meals and snacks with the app was detailed and whether it required a lot of input. In all the following sections of the paper, the complexity of the EMA protocols refers these ratings.

Active information processing

Active information processing was assessed with 7 items using the same 5-point Likert scale (1-strongly disagree to 5- strongly agree), which were developed by the study team, which comprised of experts in the study of the psychology of eating, including underlying cognitive processes. For instance, the participants rated whether recording a meal/snack with the app required them to think about what they were about to eat, or required them to know exactly how much they were about to eat.

Eating-related cognition, intentions and behaviours

Eating-related cognitions were assessed with 6 items. For example, participants indicated whether using the study app made them think about the healthiness of the meal, or become aware of how much they ate. Intentions to change behaviour was assessed with 5 items. For instance, participants rated whether using the study app made them want to eat less, or want to eat fewer snacks. Anticipated changes in eating behaviour was assessed through 5 items. Amongst others, participants rated whether using the study app would make them eat less or eat healthier. The same 5-point Likert scale was used in these three measures (1- strongly disagree to 5- strongly agree). Although this particular measure was self-developed to ensure maximum fit for the research question as well as brevity, the items were inspired by items used in the eating behaviour literature, which frequently assessed eating-related cognitions (e.g., regarding when, where and how participants consumed food; Wansink & Sobal, 2006), intentions (e.g., Reichenberger et al., 2019), and behaviours (e.g., meal skipping, Pendergast et al., 2016).



User Experience Questionnaire (UEQ)

The short form of the user experience questionnaire (Schrepp et al., 2017) was used to assess the usability of the EMA protocols. The short form of the questionnaire (8-items) was developed based on the original long version of the UEQ (26-items) and has been translated in 19-different languages (see Schrepp et al., 2017 for development, validation and reliability performance). Participants were presented with 8 semantic differentials (4 pragmatic and 4 hedonic quality items) such as complicated-easy; confusing-clear. They were asked to indicate their response using a 7-point Likert scale from 1= negative connotation to 7= positive connotation. An overall mean of the scale was calculated; higher values represent higher usability of the EMA protocols.

Experience with using nutrition app

Participants' previous experience in using nutrition apps was assessed using a stage model approach as presented in by König et.al. (2018). When they indicated that they had used an app before or were currently using an app to track their food intake, they were asked to state the app's name in an open text field.

Affinity to Technology (ATI)

ATI is a scale used to measure an individual's general comfort and willingness to engage with technology. Participants indicated their degree of interaction with technical systems (i.e., apps, software, mobile phones, pc) by indicating agreement with the four statements in the ATI (Wessel et al., 2019) scale short form. The reliability and validity of the scale has been confirmed (Franke et al., 2019; Lezhnina & Kismihók, 2020; Heilala et al., 2022) and the scale has been translated into different languages (University of Lübeck, 2025). Due to a programming mistake, responses were assessed on a 5-point instead of a 6-point Likert scale (1= completely disagree; 5= completely agree). The responses for the two negatively worded items (item 3 and 4) have been reversed before calculating the mean of the items; higher values represent higher affinity to technology.

Statistical analysis

Analyses were conducted using IBM SPSS Statistic software version 29.0.0.0. Missing values were deleted case-wise. Missing values ranged from 0% for active information processing, eating related cognition, intention and behaviour to 2,1% for perceived complexity; 2,4/ for affinity to technology and 3,2% for previous experience with nutrition apps. The main research questions (H1-3) were investigated by conducting one-way between-subject ANOVAs using the eight EMA protocols as the independent variable and the respective sub-scales for complexity, active information processing, eating-related cognitions, intentions and behaviour as the dependent variable.

The first exploratory question (E1) was tested using a one-way between-subjects ANOVA, with the eight EMA protocols as the independent variable and the UEQ score as the dependent variable. The second exploratory question (E2) was tested using an 8 (EMA protocols) x 5 (nutrition app use stage) between-subjects ANOVA with the respective sub-scales for active information processing, eating-related cognitions, intentions, and eating behaviour change as the dependent variable; since a moderating effect was tested, the report focuses on the interaction of the two independent variables. The third exploratory question (E3) was tested using a between-subjects ANCOVA, the ATI score as a covariate and the respective sub-scales for active information processing, eating-related cognitions, intentions, and eating behaviour change as the dependent variable.

Across analyses, $p = .05$ was used as the inference criterion. Reliability of scales assessing active information processing and Affinity to Technology (ATI) was calculated as internal consistency (Cronbach's alpha).

3.4. Results

Reliability analysis

Items reflecting the active information processing (Cronbach´s $\alpha = .82$), eating related cognitions (Cronbach´s $\alpha = .83$), intentions (Cronbach´s $\alpha = .80$), behaviour (Cronbach´s $\alpha = .83$) and the ATI scale (Cronbach's $\alpha = .74$) showed good internal consistencies with values above .7 (Tavakol & Dennick, 2011).

Inferential statistics results

H1: EMA protocols with more components are perceived to be more complex (manipulation check)

In line with the initial assumption, a between-subjects ANOVA ($n = 365$) revealed significant differences in perceived complexity between EMA protocols. $F(7, 365) = 16.82$, $p < .001$, $\eta^2_p = 0.24$. More specifically, EMA protocols that contained more tracking features were perceived to be more complex (perceived complexity – rated by the participants) (see Table 3 for means and standard deviations). Protocol 3, which included only the photo feature, was rated as significantly less complex than all other protocols (ps $\leq .030$), except protocol 2, which included only the classification feature. Despite also only including one tracking feature (text-based description), protocol 1 was perceived as significantly more complex than protocols 2 and 3 (ps $< .001$), but less complex than protocols 4 to 7 which contained 2-3 tracking features (ps $\leq .035$); however, there was no significant difference between protocols 1 and 8, which included four tracking features. Protocol 7, which contained three features, was rated as significantly more complex than all other protocols (ps $\leq .005$). H1 thus was largely supported. A list of all paired comparisons can be found in the online supplement.

Table 3. Complexity rating, Active Information Processing ratings, anticipated changes to eating-related cognitions, intentions, and behaviours, and usability ratings (UEQ) for the eight EMA protocols.

EMA Protocol	N	Complexity		AIP		Cognitions		Intentions		Behaviours		UEQ	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1	51	2.99	0.69	3.45	0.85	3.80	0.61	3.29	0.81	3.10	0.81	3.01	0.84
2	56	2.30	0.68	3.36	0.73	3.76	0.58	3.24	0.69	3.21	0.77	3.29	0.64
3	53	2.25	0.75	3.00	0.92	3.73	0.77	3.17	0.69	3.03	0.76	3.50	0.52
4	42	2.69	0.61	3.62	0.63	3.91	0.59	3.17	0.72	3.10	0.79	3.19	0.67
5	36	2.66	0.73	3.40	0.84	3.95	0.73	3.28	0.72	3.16	0.66	3.40	0.74
6	42	2.56	0.74	3.39	0.81	3.80	0.79	3.24	0.82	3.05	0.78	3.49	0.75
7	43	3.40	0.64	3.83	0.55	3.99	0.63	3.19	0.88	3.15	0.89	2.94	0.67
8	50	3.19	0.70	3.70	0.55	4.00	0.66	3.21	0.77	3.18	0.79	3.23	0.70

H2: The more complex the assessment tool, the more active information processing is required.

A between-subjects ANOVA (n = 373) revealed differences in active information processing between EMA protocols, $F(7, 373) = 5.63$ $p < .001$ $\eta^2_p = .097$. As expected, EMA protocols that were perceived to be more complex were also perceived to induce more active information processing, such as EMA protocol 7 (M = 3.83, SD = 0.55), compared to less complex EMA protocols, such as EMA protocol 3, which was rated as inducing significantly less AIP than all other protocols (M = 3.00, SD = 0.92; ps $\leq .012$; see Table 3 for means and standard deviations of all groups and the online supplement for the paired comparisons). H2 was thus largely supported.

H3: The more complex the assessment tool, the stronger its potential impact on participants' eating-related cognitions, intentions, and behaviour change.

Between-subjects ANOVAs (n = 373) indicate no significant differences between EMA protocols regarding perceived impact on eating-related cognitions, intentions, and behaviour (cognitions: $F(7, 373) = 1.26$, $p = .270$, $\eta^2_p = .02$; intentions: $F(7, 373) = 0.16$, $p = .993$, $\eta^2_p = .01$; behaviour: $F(7, 373) = 0.33$, $p = .942$, $\eta^2_p = .01$; see Table 3 for means and standard deviations for all groups). H3 thus was not supported.

Exploratory analysis results

E1: Do the assessment tools differ in usability?

A between-subjects ANOVA (n = 373) revealed significant differences in usability between the different EMA protocols, $F(7, 1) = 4.14$, $p < .001$, $\eta^2_p = .07$ (see Table 3 for means and standard deviations for all groups). The mock-ups that were perceived to be less complex, such as protocol 3 ($M = 3.50$, $SD = 0.52$), were also perceived as more user-friendly in comparison to protocols perceived as more complex, such as protocol 8 ($M = 3.23$, $SD = 0.67$). All paired comparisons are listed in the online supplement.

E2: Does previous experience with nutrition apps influence participants' ratings of the hypothetical assessment tools?

Two-way between-subjects ANOVAs (n = 361) indicated that differences between EMA protocols in active information processing ($F(28, 333) = 1.30$, $p = .144$, $\eta^2_p = .10$), eating related cognitions ($F(28, 333) =.58$, $p = .960$, $\eta^2_p = .05$), intentions ($F(28, 333) = 1.25$, $p = .183$, $\eta^2_p = .10$), or behaviour ($F(28, 333) = 1.15$, $p = .281$, $\eta^2_p = .09$) did not vary by participants' prior experience with using nutrition apps.

E3: Does affinity to technology influence participants' ratings of the hypothetical assessment tools?

One-way between-subjects ANCOVAs (n = 364), using the ATI score as a covariate, indicated a significant main effect of EMA protocol for active information processing ($F(1, 364) = 6.01$, $p = .015$, $\eta^2_p = .02$). Protocol 3 was rated to induce significantly less AIP compared to protocols 1, 5, 7 and 8 $ps \leq .046$; means and standard deviations and p-values for all paired comparisons are listed in the online supplement). The remaining paired comparisons did not reach statistical significance. However, there were no significant differences regarding eating-related cognitions ($F(1, 364) = 2.83$, $p = .094$, $\eta^2_p = .01$), intentions ($F(1, 364) = 1.79$, $p = .182$, $\eta^2_p = .01$), and behaviour ($F(1, 364) = 3.54$, $p = .061$, $\eta^2_p = .01$) when the ATI score was added as a covariate.

3.5. Discussion

A diverse range of smartphone-based dietary assessment tools has been used in the literature (König, van Emmenis et al., 2022), yet potential effects of this variation on comparability of findings across studies remained mostly unexplored. This study compared the perceived complexity of different smartphone-based dietary assessment tools and tested for potential differences in active information processing and anticipated impact on eating-related cognitions, intentions, and behaviours. Results indicated that EMA protocols that use more dietary tracking features were indeed perceived to be more complex and less usable, and to induce more active information processing in comparison to EMA protocols with fewer dietary tracking features. At the same time, however, the complexity of the EMA protocols was not perceived to differentially influence eating-related cognitions, intentions, and behaviours. Finally, previous experience with nutrition apps did not seem to have an impact on any of the investigated aspects of behaviour. Results were also mostly unchanged when affinity to technology was included as a covariate.

The results of this study confirmed our notion that more assessment features increase the perceived complexity of a smartphone-based dietary assessment tool; this is in line with previous research that has shown that the number of features or components in a task can influence participants' perception of its complexity (Heron & Smyth, 2010; Goldschmidt et al., 2014). Researchers might want to include more assessment features to gather more detailed information, especially when studying complex behaviours such as eating; yet our findings highlight the risk of both increasing perceived complexity and decreasing perceived usability, which might lead to underreporting, especially of snacks (Ziesemer et al., 2020), and dropout (König et al., 2021) and so reduce data quality. However, it is important to question whether recording tools indeed differentially influence compliance and dropout in EMA studies as the method is generally considered burdensome due to the number of assessments and disruptions to daily routines (Smyth et al., 2021). This assumption thus needs to be tested in future research to derive recommendations for researchers intending to use dietary EMA. Generally, however, it can be advised to select the number and type of recording features carefully so as not to overburden participants, and to pay attention to usability, especially when using food databases (König et al., 2021).

Prior research indicates that the use of food databases and requiring participants to estimate portion sizes may be more burdensome than taking pictures of the food, which translates to willingness to use the tool (Höchsmann et al., 2021). Also in the present study, EMA protocols involving food photos were seen as less complex and more usable. It could thus be recommended to use photo-based recordings instead of relying on food databases to increase compliance, although this has important implications regarding quantifying the recorded food intake (see König, van Emmenis et al., 2021, for a discussion). It is important to note that the present study investigated dietary EMA in the context of assessing – not deliberately changing – a behaviour to describe it. In this context, changes are often undesired (c.f., measurement reactivity). However, self-monitoring of (dietary) behaviour is also a frequently used behaviour change technique, which is often associated with greater intervention success (Michie et al., 2009). Thus, for the context of interventions, the present study provides different recommendations. Indeed, in line with Turner-McGrievy et al.

(2021), it could be suggested that if behaviour change is the goal, burden associated with recording might actually produce stronger intervention effects and is therefore desired.

In line with H2, we also found that the complexity of the different EMA protocols was related to the perceived level of active information processing they required. For instance, the EMA protocols that used a combination of tracking features (i.e., food database, detailed description, and portion size assessment) were perceived to require more active information processing in comparison to, for example, the EMA protocol that used only photo-based assessment. This aligns with the concept of cognitive load, where tasks perceived as more complex or demanding require higher levels of cognitive effort (Davis, 1989). Moreover, in more complex dietary-intake smartphone tracking apps, the interaction time with the tool is longer compared to less complex apps. Considering that participants in EMA studies typically track their behaviour repeatedly, this leads to an even larger amount of time spent actively processing diet-related information as the study goes on.

Based on Control Theory (Carver & Scheier, 1982), it could thus be expected that these opportunities to reflect and detect discrepancies between the current and the desired behaviour should induce changes in behaviour. Participants indeed indicated for all eight protocols that they might expect slight changes. This is also in line with prior qualitative work suggesting that EMA study participants feel that their behaviour may have changed due to taking part in a study (Eisele et al., 2022). However, contrary to our expectation, the protocols did not differ in this regard. Exploratory analyses also indicate that the ratings were not influenced by prior experience; participants who used nutrition apps before did not expect the potential influences of the EMA protocols to be different from participants without prior experience. Still, it is important to note that this study was based on a hypothetical scenario, since participants only reflected on the mock-ups and did not have the opportunity to use them. The protocols would need to be implemented and used to collect data in real life to test whether there are indeed no differences between the protocols, or whether these changes occur outside the individuals’ awareness (Eisele et al., 2023). Shedding more light on the issue of changes in behaviour due to being measured in research, i.e. measurement reactivity (French & Sutton, 2010), is especially important for dietary EMA, since research in this domain is currently lacking (König, Allmeta, et al., 2022).

Study limitations

The study has several strengths, including that it was preregistered, and materials and data are shared on the OSF. The sample size was determined a-priori and relatively large, yet due to lack of prior research the effect size estimate was based on the mean effect size for psychological studies (Open Science Collaboration, 2015). The procedure was highly standardized, with all EMA protocols using the same design and consisting of the same basic features, so differences between protocols can indeed be attributed to the combination of features and not, for instance, the visual design (Valenčič et al., 2023). However, several limitations need to be noted. First, and most importantly, the study was hypothetical and relied on self-reported expected behavioural changes which is also a limitation, given numerous accounts of overreporting of desirable and underreporting of undesirable behaviours (Paulhus, 1991; Stone, 2007). Thus, results need to be replicated in real life contexts; this would require programming apps that are identical in layout but differ in the type and number of implemented features. Unfortunately, given that the present study was conducted during the height of the Covid-19 pandemic, this was not possible at the time. Still, also self-reported information provides valuable insights, especially when later contrasted with objectively assessed data, since there is an ongoing debate in the literature as to when individuals notice changes in their behaviour (see e.g. Szymczak et al., 2020, and whether study participants are aware of changes to their behaviour due to the study context (Eisele et al., 2022). Furthermore, most constructs were assessed with items newly developed for this project, that require further validation if used again. Second, even though aspects such as affinity to technology, previous experience in using nutrition apps and the general usability (UEQ) of the EMA protocols were assessed, other aspects that might have been relevant i.e., willingness to take part in a study using EMA protocols, were not included. Third, the study sample was mostly female, young and well educated; also, former users of nutrition apps were likely overrepresented (König et al., 2018). We furthermore collected data internationally and managed recruitment from Germany; the fact that most participants thus were probably not English native speakers may have somewhat influenced their understanding of the mock-ups, descriptions and items. Yet, the sample was relatively young and thus was likely exposed to English regularly, both in school and through social and other media. The influence of the language on study outcomes is thus likely negligible. Still, generalizability to other samples thus remains to be tested. This sample composition, however, is typical for (dietary) EMA studies (see e.g., König, van Emmenis, et al., 2022, for an overview), so the conclusions drawn from this study still apply to many published studies. Finally, the sample was too small to detect small effects at sufficient power, thus small differences between EMA protocols regarding anticipated changes in eating-related cognitions, intentions and behaviours might have been missed.

3.6. Conclusion

Smartphone-based dietary EMA is becoming increasingly common in behavioural research, but apps are often developed by research teams themselves due to lack of funds for commercial solutions or specific requirements of the research project. Aside from the influence on what information is collected, the type of tracking features may influence compliance with the study protocol, which in turn influences completeness and informative value of the data. Researchers thus need to carefully select the appropriate EMA protocol for their study to balance the need for collected information with the need for high compliance.

4. Easy or difficult? Investigating perceived ease of changing eating and physical activity behaviors

Anila Allmeta [1,2], Danielle Arigo[3], Laura M. König [1,4]

[1] Faculty of Life Sciences: Food, Nutrition and Health, University of Bayreuth, Germany
[2] Department Health and Prevention, Institute of Psychology, University of Greifswald, Germany
[3] Department of Psychology, Rowan University, USA
[4] Faculty of Psychology, University of Vienna, Austria

This chapter has been published in:

Allmeta, A., Arigo, D., & König, M. L. (2026). Easy or difficult? Investigating perceived ease of changing eating and physical activity behaviors. *Applied Psychology Health and Well-Being*, *18*(1), e70124. https://doi.org/10.1111/aphw.70124

Contribution:

AA: conceptualization, methodology, formal analysis, investigation, data curation, writing-original draft; DA: conceptualization, writing-review & editing; LK: conceptualization, methodology, writing-review & editing, supervision.

4.1. Abstract

Objectives: Many people aim to eat healthier or become more physically active, yet often fail. Identifying aspects of behavior that are easier to change is crucial for effective interventions.

Methods: Two preregistered online studies assessed participants' perceived ease of changing eating and physical activity (PA) behavior and explored potential moderators.

Results: Young adults predominantly without (Study 1, N = 435, Mage = 31.6) and older adults predominantly with chronic conditions (Study 2, N = 637, Mage = 57.2) indicated the perceived ease of changing 21 aspects of eating and PA, medical history, social comparison, prior behavior change attempts and current behavior. Young adults found increasing consumption and engaging in high-intensity PA easiest and eating less and spending less time sitting most difficult to change. Older adults found reducing consumption and sedentary behavior easiest, and increasing consumption and walking at least 10.000 steps most difficult to change. Lower unhealthy food consumption correlated with easier reduction, while high PA did not always translate to perceiving further increases in PA easier. Results regarding prior attempts and social comparison were mixed across behaviors and samples.

Conclusion: Tailored intervention design should integrate users’ perceived easiness to change relevant aspects of the target behavior, which may change based on age.

4.2. Introduction

Many people aim to eat healthier (Ruf & König, 2025) or become more physically active (Oscarsson et al., 2020), yet attempts to change behavior often fail (Conner, & Sparks, 2005). Behavior change theories help understanding and facilitate lifestyle modifications, including eating and physical activity (PA). The Health Action Process Approach (HAPA; Schwarzer, 2008), for example, postulates that action and coping planning facilitate putting intentions into action by specifying and linking characteristics of a situation (when and where) and a concrete action sequence (how); the more specific, the more effective these plans should be (Schwarzer, 2008). Goal Setting Theory (GST) (Latham & Locke, 1991) also emphasizes the importance of establishing clear, measurable targets to enhance motivation and commitment towards desired behaviors.

According to GST, to be most effective, goals should be challenging yet achievable (Latham & Locke, 1991). Indeed, it could be hypothesized that some aspects of eating behavior and PA might be easier to change than others, e.g., because they require fewer cognitive resources or time and are easier to integrate into existing routines (Cheval & Boisgontier, 2021; König et al., 2022). For instance, it might be easier to increase the number of steps by getting off the bus one stop early than to make time to schedule a workout session. Steps and activity counts may be more easily modifiable than moderate to vigorous PA since they require less effort and are easier to integrate in existing daily life activities (Cheval & Boisgontier, 2021). Also, it might be easier to introduce healthy eating routines i.e., eating an apple during the coffee break, or adding a side of vegetables to a lunch instead of just stating the intention to eat healthier overall (Fogg, 2011). This makes it easier to integrate the changes into daily routines as they require less time, money, cognitive and physical resources to implement, and consequently, perceived ability increases (König & Renner, 2019).

Similarly, Fogg (2009), in his Food Behavior Model, emphasizes that the likelihood of successful behavior change can be increased by reducing its complexity and difficulty. He argues that behavior change is more likely to occur if behavior is, for instance, broken down into “baby steps” (i.e., by dividing a demanding behavior such as healthy eating into smaller actions that are more easily accomplished; c.f. “tiny habits”) to increase perceived motivation and ability to achieve the desired change.

Building on these theoretical considerations, it seems prudent to plan behavioral interventions accordingly by either offering a range of goals that most people perceive as attainable, or by steering intervention users towards behaviors that they feel they can actually reach. For this, it is useful to know the perceived difficulty of different aspects of the target behavior; critically, however, information on this is currently lacking. To address this gap, we present results from two studies that systematically assessed participants’ perception of the ease of changing different aspects of eating behavior and PA. Kaushal and Rhodes (2015) operationalized behavior complexity as participants’ perceived difficulty (and self-efficacy) in engaging in the behavior, including how hard it is to do, how much effort it takes, and whether they have the physical ability to do the behavior. In line with this, we conceptualized “ease” as individuals’ composite perception, which may be rooted in psychological factors including self-efficacy (Bandura, 1997), cognitive load (Ajzen, 1991) or contextual factors such as access to facilities.

Since it can be assumed that behaviors are perceived to be easier to change if participants already engage in them to some extent (Ouellette & Wood, 1998), we also took previous behavior change attempts and habitual eating behavior and PA into account. Additionally, we present the relationship with medical history, since chronic conditions often require individuals to modify their self-perceptions and, depending on the condition, limit or eliminate usual functions because of the symptoms or consequences of treatment (Arigo et al., 2014).

Finally, we examined the role of social comparison, or evaluating one's standing relative to others (Festinger, 1954). Comparison plays a key role in forming self-evaluations of behavioural abilities and performance (Collins, 1996; Dakin & Arrowood, 1981; Ruble et al., 1980) and has important implications for well-being and health (Gerber et al., 2018). This has been demonstrated with respect to physical health outcomes such as weight loss (Leahey et al., 2011) and in chronic illness self-care (Arigo et al., 2014). Comparison is also a core mechanism of action in health behaviour change interventions such as those to improve physical activity and eating behaviour (Olander et al., 2013). Interventions often prompt comparisons between participants to provide information about one's relative performance with respect to the target health behaviours; this is expected to trigger re-evaluation of one's performance and bolster motivation for increasing healthy behaviours (Arigo et al., 2020).

For example, when offered the opportunity to select a peer profile to read, insufficiently active women who selected "highly active" peers were often surprised that these peers' activity was similar to their own; they described re-evaluating their own activity to be better than their prior self-perceptions and a resulting increase in motivation to engage in physical activity, as increasing seemed easier after exposure to a successful peer (Arigo et al., 2025). Given the relevance of comparison for evaluating one's abilities and changing health behaviours in the context of interventions, it is likely that comparisons are related to the perceived ease of changing said behaviours. As this relation has received little attention, in the present study, we explored whether comparisons specific to eating and PA were related to the perceived ease of changing aspects of these behaviors.

This research consists of two preregistered exploratory online studies. Specifically, the goal of Study 1 was to explore potential differences in the perceived ease of change of 21 specific aspects of behavior (13 eating-related and 8 PA-related) and potential associations with social comparison, previous behavior change attempts and medical history in a convenience sample of mainly well-educated young adults, most of whom did not report chronic conditions. With Study 2, we aimed to replicate and extend the findings by additionally investigating associations between the ease of changing eating behavior and PA with habitual levels of these behaviors, in a sample of middle-aged and older adults with a higher likelihood for chronic conditions.

4.3. Methods

4.3.1. Study 1

This study was preregistered on the Open Science Framework (OSF; https://osf.io/pzwb7/), and all study materials and data are available from the OSF (https://osf.io/pukdg/). The study protocol was approved by the ethics committee of the University of Bayreuth.

Sample

A power analysis in GPower 3.1 (Faul et al. 2009) indicated a sample size of $N = 435$ for an alpha level of .001 (α-level reduced due to the large number of pairwise comparisons; Faul et al. 2009) using two-tailed dependent samples t-tests and alpha level of .05 using 2x2 within subjects ANOVAs, and a small effect size (Cohen's $d = 0.2$ or partial $\eta^2 = 0.1$) (Cohen, 1992) in direct contrasts, with at least 80% power. Inclusion criteria were being at least 18 years old and having a good command of the English language. Exclusion criteria, given the nature of the survey questions, were adults with active eating disorders or disordered eating symptoms (self-identified). A total of 435 participants completed the questionnaire; they were recruited through social media and word of mouth in winter 2022. Their mean age was 31.6 years ($SD = 11.4$), with 68.5% women, 30.8% men, and 0.7% indicating another gender. Most participants identified as not Hispanic or Latino (93.6%), and 91.5% identified as White. Most of the participants held a university degree where 38.9% indicated a Bachelor's degree, and

29.9% indicated a Master's degree as their highest qualification. The mean Body-Mass Index (BMI; kg/m^2) of the participants was 23.5 (*SD* = 4.8) and 83% of the sample did not indicate any chronic condition.

Study design and procedure

A descriptive exploratory online study was conducted using the survey platform Tivian Unipark. All participants provided informed consent prior to participating. Participants did not receive any compensation or feedback. First, participants provided demographic, anthropometric, and basic medical information. Then they were asked to compare their eating and PA behaviors to an average person similar to their demographic characteristics and living conditions, to indicate how often they thought about their eating behavior and PA, as well as to indicate previous attempts to change these behaviors. Finally, the participants were asked to indicate how easy or difficult it would be for them to change a total of 21 specific behaviors related to eating and PA.

Materials and measures

Demographic, anthropometric and medical history information.

Participants were asked to indicate their age (in years), gender (male/ female/ other) and their education level (i.e., some high school, to 8th grade, some college, no degree, Bachelor´s degree, Master´s degree). They were also asked to indicate their ethnic or racial heritage and to specify their race. Also, anthropometric information, i.e. height (in meters) and weight (in kilograms/ pounds), based on which we calculated the BMI (kg/m^2). Additionally, participants reported basic medical information based on which they were later categorized into *without* (if they reported no chronic conditions) vs *with chronic conditions* (if they reported at least one chronic condition). The list included the following conditions: arthritis, asthma, congestive heart failure, COPD (lung disease), diabetes mellitus, hyperlipidemia (high cholesterol), hypertension (high blood pressure), hypothyroidism, myocardial infarction (heart attack), nerve/ muscle disease, osteoporosis, stroke.

Social comparison. Social comparison regarding eating behavior was assessed with 2 items using 5-point scales (1 much less healthy to 5 much healthier and 1 never to 5 very frequently). Participants rated whether relative to the average person, how healthy they thought their eating behaviors were and how often they thought about their own eating behaviors relative to others'. Social comparison regarding PA was assessed with the same two 2 items referring to PA. These items were based on those used in Arigo et al. (2023) and modified to address the context of the present study.

Previous attempt to change eating and PA behaviors. The attempt to change eating behaviors, independent of its success, was assessed with 4 items. Participants were asked to indicate the number of times (in the past year) that they attempted to improve their eating behavior, started a formal eating program, started a healthy eating program on their own that lasted 3 days or less, or that lasted more than 3 days, in the past year in an open text field. The attempts to change PA levels were assessed with the same 4 items referring to PA.

Perceived ease of changing eating and PA behaviors. The perceived ease of change of 21 aspects of behavior was assessed (for mean and standard errors of the mean see Figure 1- Eating-related aspects and Figure 2- Physical activity aspects; Table A. in supplementary materials for all items means and standard deviations). These included both aspects of behavior that are usually considered "desired changes" i.e., *eat healthier, eat fewer snacks, be more physically active, spend less time sitting*, as well as aspects that are usually considered "undesired changes" i.e., *eat less healthy, skip meals, be less physically active, spend more time sitting*, to be able to paint a full picture of behavioral changes in both "healthy" and "unhealthy" directions. Additionally, 5 out of the 21 aspects of behavior ("*eating 5 portions of fruits and vegetables per day"* (WHO, 2003)*, "eating no more than 300 to 600 grams of meat* *per week"* (Clinton et al., 2020)*, "eating no more than 50 grams of added sugar per day"* (WHO, 2015), *"walking at least 10.000 steps per day"* (CQ University's Physical Activity Research Group, 2001)*, "engaging in moderate to high intensity activities for at least 300 minutes per week"* (Bull et al., 2020), were based on international recommendations and guidelines. The terms added sugar and moderate to high intensity activities were further defined in a footnote; see the materials on the OSF.

Participants were asked to rate the 21 aspects on a six-point scale from (1) would be very difficult for me to (6) would be very easy for me. When rating, they were explicitly asked to compare each behavior listed to what they were currently doing.

Figure 1. Means and standard errors of the mean of perceived ease of changing eating behaviors, presented separately for the samples in study 1 and study 2.

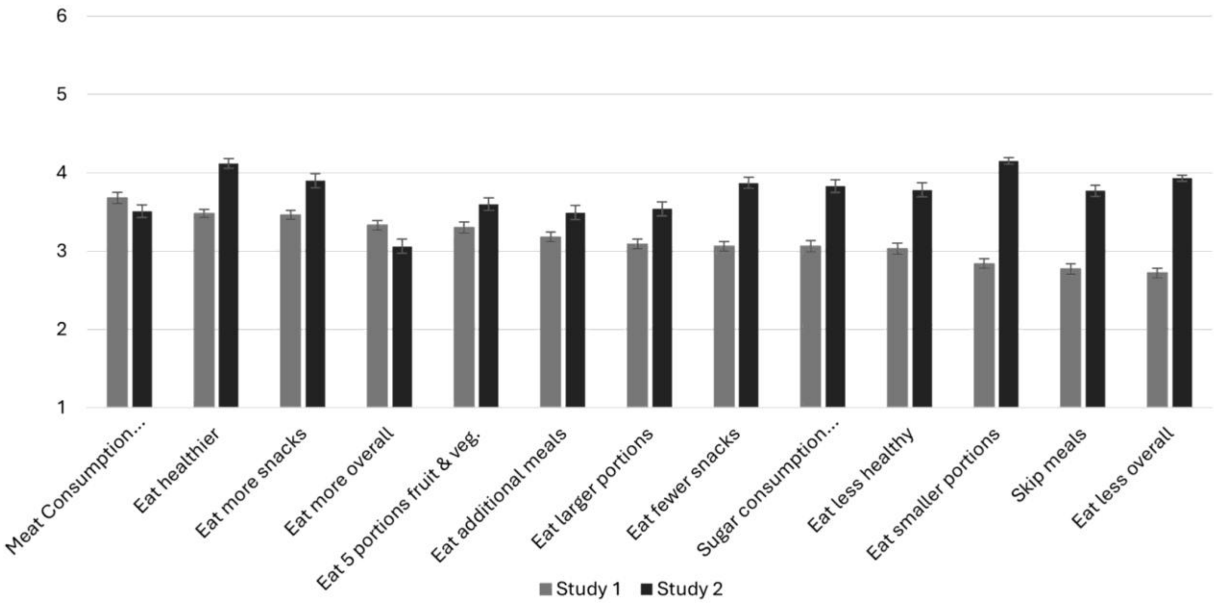
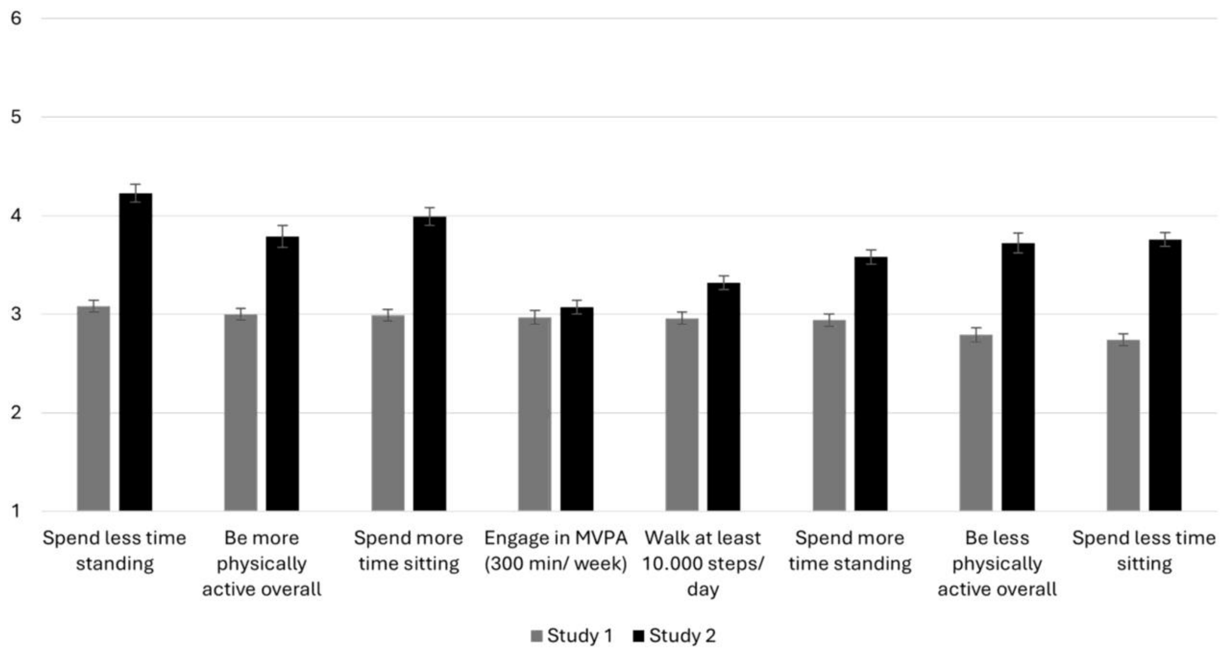


Figure 2. Means and standard errors of the mean of perceived ease of changing physical activity behaviors, presented separately for the samples in study 1 and study 2.



Statistical Analysis

Analyses were conducted using IBM SPSS Statistics software version 30.0.0. All participants completed the study questionnaire and there were no missing values for the outcomes of interest. The 13 aspects of eating behavior and 8 aspects of PA were compared using within-subject ANOVAs. Exploratory pairwise comparisons between each combination of items were conducted to follow up significant main effects. Pearson correlations between the 21 aspects and social comparison and the number of previous behavior change attempts were conducted. Diverging from the pre-registration, independent samples t-tests were conducted to compare perceived ease of changing aspects of eating and PA behaviors between participants with and without chronic conditions. Results were considered statistically significant at $\alpha = 0.05$ for the omnibus test and at $\alpha = 0.001$ for the pairwise comparisons (due to the large number of relevant comparisons). Effect sizes, respective to the statistical tests conducted i.e., Cohen´s *d*; Pearson´s *r*, were interpreted according to Cohen (1992).

4.3.2. Study 2

This analysis used data from a larger study on cognitive decline (Jeong et al., 2025). This analysis was preregistered on the Open Science Framework (OSF; https://osf.io/x7dyb/), and all study materials and data relevant to this analysis are available from the OSF (https://osf.io/pukdg/). The study protocol was approved by the Institutional Review Board at Clemson University.

Sample

Since the research questions addressed in this study were not the primary aim of this data collection, no formal sample size planning for this study was conducted. However, a power analysis in GPower 3.1 (Faul et al., 2009) confirmed that the target sample size of N = 400[1] for the larger project would be sufficient to detect small effects (Cohen's d = 0.2 and partial $\eta^2 = 0.1$) in direct contrasts using two-tailed dependent samples t-tests for an alpha level of .001 (α-level reduced due to the large number of pairwise comparisons (Faul et al., 2009) and in contrast using 2x2 within subjects ANOVAs; with at least 80% power. Inclusion criteria were being at least 40 years old[2], having a good command of the English language, living independently and not being diagnosed with dementia or other cognitive impairment.

A total of 658 participants completed the questionnaire; they were recruited online via the Qualtrics panes in the USA in spring and summer 2023. In the final data analysis, 637 participants were included; 19 participants were removed since they indicated to be younger than 40 and 2 were removed because they did not answer at least 25% of the questions. Participants' mean age 57.2 years (SD = 12.6), with 49.5% women, 49.8% men, and 0.7% indicating another gender. Most of the participants (94%) identified as not Hispanic/ Latino/ Spanish and the majority (79.6%) identified as White; 22.8% held a Bachelor's degree, and

[1] In the preregistration, this was mistakenly noted as N = 500

[2] In the preregistration, this was mistakenly noted as aged 50 and older.

12.6% a Master's degree. The majority of participants (82.4%) indicated having at least one of the chronic medical conditions listed in the questionnaire.

Study design and procedure

Data were collected online as part of a larger study on cognitive decline and social comparison in older adults conducted via Qualtrics. All participants provided informed consent prior to participating in the study. Participants were compensated based on the agreement they completed when joining the Qualtrics survey panel. For example, some participants were compensated with travel rewards (e.g., airline miles) if they joined Qualtrics through their travel rewards program.

Participants provided demographic, anthropometric, self-rated health and medical history information. Afterwards, participants completed questionnaires on personality traits, mental health, life satisfaction, cognition, and coping. Then they were asked to compare their eating and PA behaviors to average person similar to their demographic characteristics and living conditions, to indicate how often they thought about their eating behavior and PA, as well as to indicate previous attempts to change these behaviors. Consequently, the participants were asked to indicate how easy or difficult it would be for them to change a total of 21 specific behaviors related to eating and PA. Finally, they were asked to complete the short version of the International Physical Activity Questionnaire (IPAQ) (Craig et al., 2003) and a Food Frequency Questionnaire (FFQ) (Winkler & Döring, 1998).

Materials and measures

Demographic, anthropometric and medical history information, social comparison regarding eating behavior and PA, previous attempts to change eating behavior and PA, and the perceived ease of changing of 21 aspects of eating behavior and PA were assessed in the same manner as in Study 1.

Eating Behavior. Participants were asked to complete the FFQ by Winkler & Döring (1998), which assesses consumption frequency of 24 food groups on a scale from 1 = almost every day to 6 = never. Based on the scoring algorithm proposed by Winkler & Döring (1998), an overall evaluation score of participants' eating behavior was calculated. In addition, for individual food groups related to individual aspects of behaviors (see Table 1), variables were grouped to represent consumption frequency in line with the behavioral aspect of interest.

Table 1.
Correspondence between: FFQ items and aspects of eating behavior. Summary of correlation analyses conducted in Study 2.

FFQ	**Aspects of eating behavior**	**r**	**df**	***p*-value**
Overall score based on Winkler & Döring (1995)	eating healthier	.26	509	<.001*
	eating less healthy	-.19	508	<.001*
Chocolate, Pralines	eating no more than 50 grams of added sugar per day	.11	505	.017*
Cakes, pastries, biscuits		.15	503	<.001*
Other sweets (candys, compote)		.20	504	<.001*
Fruit juices, other soft drinks (lemonade, cola-beverages, and others)		.05	504	.227
Diet lemonades, other diet beverages		-.07	503	.105
Meat (without sausages)	eating no more than 300 to 600 grams of meat per week	.28	501	<.001*
Sausages, ham		.16	500	<.001*
Poultry		.06	501	.164
Salad or vegetable, raw	eating five portions of fruits and vegetables per day	-.38	506	<.001*
Vegetable, cooked		-.33	507	<.001*
Fresh fruit		-.41	507	<.001*
	eating fewer snacks	.22	508	<.001*

Salted snacks, such as salted peanuts, crisps, etc.		.14	509	.002*
Chocolate, Pralines		.11	507	.010*
Cakes, pastries, biscuits				
Other sweets (candys, compote)		.10	508	.031*
Salted snacks, such as salted peanuts, crisps, etc.	eating more snacks	-.17	508	<.001*
Chocolate, Pralines		-.15	509	<.001*
Cakes, pastries, biscuits		-.16	507	<.001*
Other sweets (candys, compote)		-.21	507	<.001*

Note. FFQ = Food Frequency Questionnaire.

*Correlations significant at .005 level (2-tailed)

Physical activity. PA was assessed by asking the participants to complete the short version of the IPAQ (IPAQ-SF). Due to an error in the set-up of the survey, participants reported the activity time in ranges and not hours and minutes. Therefore, diverging from the preregistration, instead of considering the collected data as continues variables (calculating MET-minutes), the data were treated as categorical and further processed following the scoring algorithm of the IPAQ-SF (Sjostrom 2005). Participants were categorized into three levels of PA: low, moderate, and high. Walking and sitting time were considered as ordinal variables. The corresponding IPAQ scores and items and the aspects of PA assessed in the questionnaire are presented in Table 2.

Table 2.
Correspondence between: IPAQ-SF scores and aspects of physical activity. Summary of analyses conducted in Study 2.

IPAQ	**Aspects of Physical Activity**	**t**	***df***	**p-value**	**Cohen´s *d***	**Yes**		**No**	
						M	***SD***	**M**	***SD***
Low physical activity	being more physically active	3.61	508	.495	1.50	3.62	1.49	4.28	1.59

	being less physically active	-5.27	501	.279	1.51	3.96	1.49	2.97	1.63
Moderate physical activity	being more physically active	-3.69	508	365	1.50	4.30	1.60	3.61	1.48
	being less physically active	4.80	501	.161	1.52	3.04	1.67	3.94	1.49
High physical activity	being more physically active	-7.86	508	.026*	1.44	4.45	1.33	3.38	1.48
	being less physically active	4.96	501	.001*	1.52	3.31	1.65	4.03	1.45
Moderate physical level	engage in moderate to high intensity activities for at least 300 min/week	-7.29	507	.659	1.60	4.30	1.56	2.86	1.60
High physical activity level		-12.35	507	.446	1.47	4.26	1.53	2.53	1.44
		Sperman´s r*ho*				***df***		***p*-value**	
Walking time	walking at least 10000 steps/day	.41				505		<.001*	
Sitting time	spending less time sitting	-.27				498		<.001*	
	spending more time sitting	.20				500		<.001*	

Note. IPAQ-SF = International Physical Activity Questionnaire – Short Form
* significant at .005 level (2-tailed)

Statistical analysis

Analyses were conducted using IBM SPSS Statistic software version 30.0.0.0. Missing values were deleted case-wise. Missing values in Study 2, did not exceed 20% of the total number of participants (18.4%-19.3%). Nevertheless, in each statistical test, the number of participants was never below the sample size (n = 400) calculated to achieve at least 80% power and detect a small effect size (Cohen’s d= 0.2). In line with Study 1 but diverging from the preregistration, the 13 aspects of eating behavior and 8 aspects of PA were compared using within-subject ANOVAs. Exploratory pairwise comparisons between each combination of items were conducted to follow up significant main effects. Additionally, to align with the analysis of Study 1, Pearson correlations between the 21 aspects and social comparison and the number of previous behavior change attempts were conducted. The associations between perceived ease of change and the frequency of consuming certain foods and beverages as indicated in the FFQ were analyzed with Pearson correlations. Diverging from the preregistration but in line with the adjusted coding of the IPAQ, the associations between perceived ease of change and the IPAQ scores and items were tested using independent samples t-tests (for categorical variables) and Sperman rank correlations (for ordinal variables). Lastly, the association between perceived ease of change and medical history (with vs without chronic conditions) were tested using independent samples t-tests. Results were considered statistically significant at $\alpha = 0.05$ for the omnibus test and at $\alpha = 0.001$ for the pairwise comparisons (due to the large number of relevant comparisons). Results were interpreted based on the effect size, respective to the statistical tests conducted i.e., Cohen´s *d*; Pearson´s *r*, according to Cohen (1992).

4.4. Results

4.4.1. Study 1

Do aspects of eating behavior and PA differ in the perceived ease of change?

A within-subjects ANOVA yielded significant differences for aspects of eating behavior ($F(12, 423) = 21.08$, $p < .001$, partial $\eta^2 = 0.37$). Participants perceived "*eat[ing] no more than 300-600 grams of meat per week"* to be the easiest aspect to change ($M = 3.68$, $SD = 1.54$) and "*eat[ing] less overall*" to be the most difficult ($M = 2.72$, $SD = 1.20$). In Table A in the supplementary materials, means and standard deviations for each of the 13 aspects of eating behavior are reported; results of all post hoc paired comparisons are listed in the supplementary material, Table B. In brief, there were significant differences for almost all pairwise comparisons between the 13 aspects of eating behavior (p-values .049 to <.001). Aspects of eating behavior where individuals would increase consumption i.e., eat more overall, eat more snacks, eat healthier, or eat 5 portions of fruit and vegetables per day, were generally perceived as easier to change in comparison to aspects where they would need to reduce the consumption i.e., eat less overall, eat smaller portions, or skip meals.

Significant differences were also found for aspects of PA ($F(7, 428) = 5.28$, $p < .001$, partial $\eta^2 = 0.08$). Participants perceived "*spending less time standing*" (M = 3.08, SD = 1.22) to be the easiest aspect to change and "*spending less time sitting*" (M = 2.74, SD = 1.25) to be the most difficult aspect to change. In Table A (in the supplementary material), means and standard deviations for each of the 8 aspects of PA are reported; post-hoc paired comparisons are listed in the supplementary material, Table C. In brief, aspects such as being more physically active, engaging in moderate to vigorous physical activity for 300 min/week, walk at least 10.000 steps/day seem to be perceived easier to change in comparison to aspects such as spending less time sitting, spending more time standing or be less physically active overall. Mainly, there were significant differences only between the following aspects of PA: be more physically active and spend less time sitting and engage in moderate-vigorous physical activity for at least 300min/week and walk at least 10.000 steps/day; p-values ranged from .048 to <.001.

Are social comparisons associated with perceived ease of change?

Pearson correlations were conducted to examine the relationship between social comparison and the perceived ease of changing eating and PA behaviors. All results are presented in the supplementary material, Table D. In brief, small to medium significant positive correlations were found between participants' ratings on the healthiness of their eating behavior in comparison to the average person with similar characteristics as themselves and the ease of changing healthy behaviors such as "*eat[ing] smaller portions*" (r = .10, p =

.034), "*eat[ing] healthier overall*" (r = .39, *p* <.001), or "*eat[ing] fewer snacks*" (r = .20, *p*<.001), indicating that people who think that they eat healthier than others also find it easier to engage in these behaviors. There was also observed a negative significant correlation with the item "*eat less healthy* (r = -.33, *p* = <.001)", indicating that participants who think that they eat healthier than others found it more difficult to eat less healthy. In addition, there were small significant positive correlations between the how often participants thought of their eating behaviors in relative to others and the items "*eat more in general*" (r = .10, *p* = .036) and "*eat larger portions*" (r = .12, *p* = .012), indicating that more frequent social comparisons were associated with participants thinking that eating more and larger portions would be easier for them.

Regarding PA, there was a small but significant positive correlation between participants' rating of the level of their PA in comparison to the average person with similar characteristics as themselves and the item "*engage on moderate/vigorous physical activity for 300min/week*" (r = .13 *p* =.008), indicating that participants who think that they were more physically active than others found it easier to engage in recommended levels of moderate to vigorous PA. Small to medium positive significant correlations were also found between the how often participants thought of their PA levels in relative to others and the desired aspects of PA, including "*be more physical active overall*" (r = .34 *p* <.001) *; "walk 10.000 steps/day*" (r = .34, *p* <.001) and "*engage in moderate/vigorous physical activity for 300min/week*" (r = .60, *p* <.001). In addition, the results showed small to medium negative significant correlations for undesired PA aspects: "*be less physically active overall*" (r = -.37, *p* <.001); "*spend less time standing*" (r =- .14, *p* =.003) and "*spend more time sitting*" (r = -.26, *p* <.001). Taken together, these results indicate that participants who compared themselves more frequently to others regarding their PA behaviors found it easier to be more physically active. All results are presented in the supplementary material, Table E.

Are previous behavior change attempts associated with perceived ease of change?

To investigate the associations between previous attempts to change eating behavior and PA with the participants perceived ease of changing these behaviors, Pearson correlations were conducted (see the supplementary material, Table F and G respectively for all results). Small yet significant positive correlations between previous attempts to improve eating behavior and "*eat more overall*" (r = .15, *p* = .002); "*eat larger portions*" (r = .11, *p* = .021); and "*eat less healthy*" (r = .13, *p* = .008) were found, indicating that participants with more prior eating behavior change attempts found it easier to engage in undesired eating behaviors. Also, a small significant positive correlation resulted between the number of times participants started a formal eating program and "*eat additional meals*". Meanwhile, there were medium to large significant positive correlation between the number of times participants started a formal exercise program and "*walk[ing] 10.000 steps/day*" (r = .10, *p* = .049) and between previous attempts to improve PA level and "*being less physically active*" (r = .11, *p* = .036).

Is the medical history associated with perceived ease of change?

Independent samples t-tests were conducted to compare each of the 13 aspects of eating behavior between participants with and without chronic conditions. The t-tests showed significant differences between the two groups for four of the 13 eating-related aspects (see Table 3 for all results). Generally, participants without chronic conditions perceived it to be easier to "*skip meals*" and more difficult to change the other three aspects of eating behavior compared to participants with chronic conditions. In addition, the results of the t-test showed significant differences between the two groups for three of the 8 PA aspects (see Table 4 for all results). Generally, participants without chronic conditions perceived it to be easier to change their PA in comparison to participants with chronic conditions.

Table 3.
Associations between medical history and perceived ease of change in Study 1

Aspects of Behavior	t	df	*p*-value	Cohen´s d	Without chronic conditions		With chronic conditions	
					M	***SD***	**M**	***SD***
Eating Behavior:								
Eat less overall	.03	433	.979	1.20	2.72	1.19	2.72	1.29
Eat smaller portions	-.31	433	.760	1.21	2.83	1.18	2.88	1.35
Eat more overall	.90	433	.367	1.93	3.35	1.27	3.20	1.39
Eat larger portions	.25	433	.802	1.34	3.10	1.32	3.05	1.46
Eat healthier	.53	433	.598	1.10	3.49	1.10	3.42	1.08
Eat less healthy	-.30	433	.767	1.35	3.02	1.35	3.07	1.34
Skip meals	-4.04	433	<.001*	1.40	2.65	1.40	3.36	1.37
Eat additional meals	1.75	433	.081	1.33	3.23	1.31	2.93	1.43
Eat fewer snacks	1.78	433	.076	1.26	3.01	1.25	3.30	1.29
Eat more snacks	2.82	433	.005*	1.22	3.53	1.19	3.09	1.37
Eat 5 portions of fruit and vegetables	2.06	433	.040*	1.45	3.37	1.44	2.99	1.55
Eat no more than 300-600gr of meat/week	2.21	433	.027*	1.53	3.76	1.52	3.32	1.56
Eat no more than 50gr of added sugar/day	1.18	433	.237	1.36	3.10	1.35	2.89	1.40
Physical Activity:								
Be more physically active overall	2.45	433	.015*	1.18	3.06	1.18	2.69	1.19
Be less physically active overall	.23	433	.816	1.19	2.80	1.38	2.76	1.39
Spend more time standing	1.07	433	.285	1.18	2.97	1.15	2.81	1.33
Spend less time standing	-.35	433	.724	1.22	3.07	1.19	3.12	1.37
Spend less time sitting	.89	433	.371	1.52	2.76	1.25	2.62	1.27
Spend more time sitting	-1.15	433	.252	1.28	2.96	1.26	3.15	1.36
Walk at least 10.000 steps/day	2.29	433	.023*	1.33	3.02	1.31	2.64	1.38
Engage in moderate/vigorous physical activity for 300min/week	2.73	433	.007*	1.34	3.05	1.32	2.58	1.41

Note. *significant at .005 level (2-tailed)

4.4.2. Study 2

Do aspects of eating behavior and PA differ in the perceived ease of change?

A within-subjects ANOVA yielded significant differences for aspects of eating behavior ($F(12, 5868) = 12.38$, $p < .001$, partial $\eta^2 = .025$; see Table A (in supplementary material). More specifically, participants perceived "*eat[ing] smaller portions"* to be easiest to change ($M = 4.15$, $SD = 1.29$) and "*eat[ing] additional meal*" to be most difficult to change ($M = 3.49$, $SD = 1.51$). In general, desirable aspects of eating behavior such as eating smaller portions, eating healthier or eat less overall were perceived as easier to change in comparison to undesirable aspects such as eating additional meals or eating larger portions. Post-hoc tests revealed differences in almost all pairwise comparisons between the 13 aspects of eating related behavior (p-values .040 to <.001; see supplementary material, Table H).

Significant differences were also found for aspects of PA ($F(7, 3465) = 34.67$, $p < .001$, partial $\eta^2 = .065$; see Table A (in supplementary material). Participants perceived "*spending less time standing*" (M=4.23, SD=1.34) to be the easiest to change and "*engage[ing] in at least 300 minutes of moderate to vigorous physical activity per week*" (M=3.07, SD=1.68) to be the most difficult. In brief, general aspects of behavior such as being overall more physically active, spending less time standing or more time sitting were perceived to be easier to change in comparison to very specific aspects of behavior such as engaging in moderate-vigorous physical activity for 300min/week or walking at least 10.000 steps/day. Post-hoc tests revealed significant differences in almost all paired comparisons (p-values between .014 and <.001; see supplementary material, Table I).

Are social comparisons associated with perceived ease of change?

Although not planned in the pre-registration, we aligned Study 2 with Study 1 and also investigated whether social comparisons were associated with perceived ease of change. All Pearson correlations and specific descriptions are presented in the supplementary material (Table J and K for PA and eating behaviour respectively). In brief, small to medium significant positive correlations were found between participants' ratings of the healthiness of their eating behavior in comparison to the average person with similar characteristics as themselves and the ease of changing healthy behaviors such as eating less overall, eating healthier overall, or eat 5 portions of fruit and vegetables, indicating that participants who think that they eat healthier than others also find it easier to engage in these behaviors. However, there were no significant correlations between the how often participants thought of their eating behaviors in relative to others. Regarding PA, participants who compared their PA levels more frequently to others, perceived it easier to be more physically active.

Are previous behavior change attempts associated with perceived ease of change?

Again, to align Study 2 with Study 1, we also investigated whether previous behavior change attempts, were associated with perceived ease of change. All results of the analysis (Pearson´s correlations) and specific descriptions are presented in the supplementary material (Table L and M for eating and PA behaviour respectively). Generally, the results indicate that that participants with more prior eating behavior change attempts found it easier to engage in desired eating behaviors. Meanwhile, regarding previous attempt to change PA levels, the results would suggest that participants that started formal exercise programs would perceive it to be more difficult to "*be less physically active*" and easier to "*engage[ing] in mod-vigorous physical activity for at least 300 min/week*".

Is the medical history associated with perceived ease of change?

Independent samples t-tests were conducted to compare each of the 13 aspects of eating behavior between participants with and without chronic conditions. The t-tests showed significant differences between the two groups for seven of the 13 eating-related aspects (see Table 4 for all results). Generally, participants with chronic conditions perceive it to be more difficult to change these aspects of eating behavior than participants without chronic conditions. In addition, the results of the t-test, showed significant differences between the two groups for five of the 8 PA aspects (see Table 4 for all results). Generally, participants with chronic conditions perceive it to be more difficult to change these aspects of PA in comparison to participants without chronic conditions.

Table 4.
Associations between medical history and perceived ease of change in Study 2.

Aspects of Behavior	t	df	*p*-value	Cohen´s d	Without chronic condition		With chronic condition	
					M	***SD***	**M**	***SD***
Eating Behavior:								
Eat less overall	-2.59	505	.010*	1.32	4.29	1.16	3.85	1.34
Eat smaller portions	-2.33	505	.020*	1.28	4.48	1.17	4.10	1.30
Eat more overall	-.62	502	.536	1.48	3.88	1.37	3.76	1.50
Eat larger portions	-1.37	504	.170	1.53	3.78	1.39	3.52	1.55
Eat healthier	-3.04	505	.003*	1.35	4.55	1.14	4.03	1.39
Eat less healthy	.37	505	.709	1.46	3.74	1.44	3.81	1.46
Skip meals	-.83	506	.406	1.55	3.92	1.61	3.75	1.55
Eat additional meals	-2.47	504	.081	1.33	3.84	1.54	3.44	1.50
Eat fewer snacks	-2.33	505	.020*	1.42	4.26	1.32	3.82	1.44
Eat more snacks	-.86	505	.391	1.46	4.03	1.33	3.87	1.48
Eat 5 portions of fruit and vegetables	-2.92	502	.004*	1.57	4.10	1.57	3.51	1.57
Eat no more than 300-600gr of meat/week	-2.37	497	.018*	1.57	3.92	1.45	3.44	1.59
Eat no more than 50gr of added sugar/day	-1.18	500	.240	1.50	4.01	1.42	3.79	1.51
Physical Activity:								
Be more physically active overall	-4.19	505	<.011*	1.49	4.38	1.31	3.59	1.52
Be less physically active overall	.74	498	.460	1.55	4.38	1.31	3.59	1.52
Spend more time standing	-5.75	501	<.001*	1.56	4.54	1.27	3.40	1.60
Spend less time standing	1.15	502	.251	1.33	4.28	1.32	4.54	1.27
Spend less time sitting	-3.33	501	<.001*	1.41	3.05	1.32	2.58	1.41
Spend more time sitting	.03	502	.979	1.41	4.01	1.40	4.02	1.41
Walk at least 10.000 steps/day	-7.40	503	<.001*	1.56	4.57	1.15	3.10	1.61
Engage in moderate/vigorous physical activity for 300min/week	-5.86	503	<.001*	1.62	4.11	1.55	2.90	1.64

Note. *significant at .005 level (2-tailed)

Is the frequency of consuming specific foods and beverages related to perceived ease of changing related aspects of eating behavior?

To examine the association between the frequency of consuming different foods and beverages, indicated on a scale from (1) nearly every day to (6) never, and the perceived ease of change of the corresponding eating related aspects, Pearson´s correlations where conducted. Results are summarized in Table 1. Participants who reported lower consumption frequency (represented by higher values on the item) also reported to be easier to change "eat[ing] healthier" and difficult to change "eat[ing] less healthy".

There were small significant positive correlations between the perceived ease of "*eat[ing] no more than 50g of added sugar per day*" and the consumption frequency of the following food items; *"chocolate, pralines"* "*cakes, pastries, biscuits*" and "*other sweets, candy, compote*"; due to the scoring of the FFQ items, this indicates that higher perceived ease of reducing sugar consumption was associated with lower consumption frequency of unhealthy foods.

Additionally, there were small positive correlations between the perceived ease of change of the meat consumption eating aspect: "*eat no more than 300-600g of meat/week"* and the consumption of the following food items: "*meat (without sausage*") and "*sausage, ham*"; due to the scoring of the FFQ items, this indicates that higher perceived ease of reducing meat consumption was associated with lower consumption frequency of meat-based foods.

Moreover, the correlation analysis revealed moderate, negative correlations between the perceived ease to change "*eat 5 portions of fruit and vegetables per day*" and the following food items: "*salad or vegetables (raw)*" "*vegetables cooked*" and *fresh fruits*"; due to the scoring of the FFQ items, this indicates that higher perceived ease of increasing fruit and vegetable consumption was associated with higher consumption frequency of fruit and vegetable-based foods.

Lastly, there were small, positive correlations between the perceived ease of change regarding snack consumption ("*eat fewer snacks*") and the consumption frequency of the following food items; "*salted snacks*"; "*chocolates, pralines*"; "*cakes, pastries, biscuits*"; and "*other sweets, candy, compote*". While there were small, negative correlations between "*eat more snacks*" and "*salted snacks*"; "*chocolates, pralines*"; "*cakes, pastries, biscuits*"; and "*other sweets, candy, compote*"; due to the scoring of the FFQ items, this indicates that higher perceived ease of reducing snack consumption was associated with lower consumption frequency of different types of snacks.

Are PA levels associated with the perceived ease of change?

First, we examined the relationship between PA levels and the perceived ease of change of the corresponding PA related aspects using independent samples t-tests. All results are summarized in Table 2. There were no significant differences in the perceived ease of change of "*engage[ing] in moderate-high intensity physical activity for at least 300 min/week"* and participants that had (or not) a moderate level or the ones that had (or not) a high level of PA.

Regarding the general PA level, independent-samples t-tests between each PA category (low, moderate and high PA level) and the perceived ease of "*being more and being less physically active*" were conducted. There were no significant differences in the perceived ease of change of "*being more physically active*" or "*less physically active*" and the participants that had (not) a low or a moderate level of PA level. Instead, there were significant differences in the perceived ease of change of "*being more physically active*" or "*being less physically active*" and the participants that had (or not) a high level of PA i.e., participants that had a higher level of PA would perceive it more difficult to further increase their PA level ("*be more physically active*"), and easier to decrease their PA level ("*be less physically active*").

Correlations were conducted for the remaining indicators for PA; since the scales were ordinal, Spearman correlations were used. Results are summarized in Table 2. There was a moderate positive correlation between the time spent walking and the perceived ease of "*walking at least 10.000 steps/day*", thus, participants that spent more time walking, perceived it to be easier to walk at least 10,000 steps a day. Furthermore, there was a small negative correlation between the sitting time and the perceived ease of "*spending less time sitting*" and a small positive correlation and the perceived ease of "*spending more time sitting*". Thus, participants who spent more time sitting perceived it to be more difficult to reduce their sitting time, and instead perceived it easier to increase the time they spent sitting.

4.5. Discussion

Health behavior theories suggest that achievable goals facilitate behavior change. This research thus was designed to assess participants' perceptions of the ease of changing a broad range of eating and physical activity behaviors to guide tailored interventions. We also explored how age, medical history, previous attempt to change behavior, social comparisons, and frequency of consumption of specific foods and beverages, physical activity, and sitting time are associated with these perceptions. Overall, different aspects of eating and PA were indeed seen as differing in ease to change. These differences were related to age and medical history, suggesting a potential influence of an individual's physical health status and associated external factors. Furthermore, at least for eating behaviors, healthier food intake was associated with finding engaging in further healthy eating behaviors easier; this is further supported by the finding that prior behavior change attempts were also associated with perceived ease. Together, these results underline the complexity, but also necessity, of tailoring interventions to individual needs to promote their effectiveness.

Overall, the observation that the participants in this study indeed viewed different aspects of eating and PA as differing in ease to change mirrors prior research. Using a similar method to our study, Dorina et al. (2024) asked participants to indicate the complexity of 24 behaviors related to health and sustainability, showing that perceived complexity is influenced by several salient characteristics and that complex behaviors (e.g., abstaining from smoking), are more likely to be true to the previously identified characteristics, compared to simple behaviors (e.g., eating fruit). Importantly, however, Dorina et al. (2024) asked participants to rate the complexity for an average person, while the present research focused on individuals' perceptions for themselves. Additionally, they focused on different behaviors, while in this study the focus is on different aspects of two behaviors (eating-related and PA behavior). The perceived difficulty or ease of changing a behavior is likely rooted in a complex interplay of psychological factors. According to Kaushal and Rhodes (2015), the perceived difficulty of a behavior is influenced by self-efficacy, with McCloskey and Johnson (2019) additionally suggest that the time, attention, and planning required to prepare and execute a behavior are also considered. Attempting to change perceived ease of behavior thus likely requires a multifaceted intervention, providing further support for the underlying assumption of the Fogg Behavior Model (Fogg et al., 2009) to adjust the intervention's target behavior instead to increase motivation.

Differences based on age and medical history

Regarding eating behavior, younger adults predominantly without chronic conditions (Study 1 sample) found increases in consumption easier, which might align with hedonic goals and social facilitation of eating, that emphasize pleasure, satiety, and shared experiences over strict dietary control (McInnes et al., 2023; Stok et al., 2016; Vartanian et al., 2007). Without immediate need for dietary restrictions, the notion of "*eating more*" can feel less effortful than "*eating less*", which would involve restraint and perceived deprivation.

Anticipated changes might thus be viewed through the lens of convenience, taste, or social acceptance rather than broad health management as could be the case for adults with chronic conditions (McInnes et al., 2023). In contrast, participants predominantly with chronic conditions (Study 2 sample) found it easier to make healthier choices i.e., *eating smaller portions* in comparison to less healthier choices i.e., *eating additional meals*, which could be driven by higher health motivation (c.f. extrinsic adaptive motivation; Ryan & Deci, 2017; Ahmadi et al., 2023) and clearer outcome expectancies directly linked to their health status (Rosenstock et al., 1988). For individuals managing chronic conditions, the link between dietary choices and disease management becomes notable and often a primary motivator. They may have developed self-efficacy through past experiences or guidance from healthcare professionals, recognizing that these changes, while requiring effort, yield considerable health benefits that are highly valued (Bandura, 1997). The awareness of the importance of healthy choices, not only for general health but also for preventing or coping with specific conditions, might reorient their perception of “ease”. What might be perceived as restrictive by a healthy young adult might become a necessary and therefore more "doable" (or even empowering) strategy for managing their health for someone with a chronic illness.

Both groups perceiving increases in PA (compared to decrease in PA) as easier to change is consistent with an overarching positive societal value placed on being active (Sallis & Owen, 1999). However, young adults (Study 1 sample) favoring concrete increases (i.e., *300 min MVPA/week, 10,000 steps/day*) versus adults with chronic conditions (Study 2 sample) preferring *less time sitting* or *more time standing* indicates how perceived capability and different barriers might influence the perception of ease For instance, for young adults, who face fewer health-related limitations, specific and challenging PA goals like achieving a certain step count or vigorous activity duration are often perceived as highly attainable (Locke & Latham, 2002). Their self-efficacy for these performance-based targets is likely robust, derived from past successes and a general expectation of physical capability. This aligns with goal-setting theory, which states that specific, difficult goals can be highly motivating – both intrinsically and extrinsically – and lead to greater effort and achievement, especially when individuals believe they have the means to attain them (Locke & Latham, 2002; Swann et al., 2025). Instead, for adults with chronic conditions (Study 2 sample), the perceived ease associated with reducing sedentary behavior rather than increasing vigorous activity could be adaptive. For this population, traditional PA goals might be impacted by symptoms such as pain, fatigue, or functional limitations, making large increases in structured exercise feel daunting and potentially risky (Rebar et al., 2016; Shields et al., 2012; Ryan & Deci, 2017; Ahmadi et al., 2023). Reducing sedentary time, while still challenging due to being strongly habitual, might still be perceived as a more manageable and less physically demanding target (Gardner et al., 2012). This reflects a pragmatic approach to perceived behavioral control, where individuals prioritize behaviors they feel more confident about given their health context.

Additionally, age and health status might also shift the behavior change approach from specific, tangible actions to more holistic, long-term lifestyle modifications. For instance, for younger adults and healthy individuals in Study 1, achieving specific, tangible goals derived from guidelines might provide immediate positive reinforcement, making those aspects seem easier and more directly impactful on their daily lives. In contrast, older adults (Study 2), especially those managing complex health conditions, might shift their focus towards maintaining health and well-being rather than achieving specific, potentially difficult, performance-oriented goals. At the same time, broad changes like "be[*ing*] more active overall" might feel especially difficult due to the expected increased cognitive load and the cumulative nature of managing multiple health aspects, making small, specific "wins" less impactful or harder to identify within a larger, more ambiguous goal (Charness, & Boot, 2016). Sawyer and McManus (2021) qualitatively explored the experience of adults with chronic conditions such as type 2 diabetes and hypertension attempting lifestyle changes including PA and dietary habits. Their results indeed showed that participants often felt overwhelmed by the gap between the current and the desired state, suggesting that broader, more general goals like “losing weight” or “eating healthier” felt daunting without breaking them down into smaller, manageable steps. Therefore, interventions must consider these distinct perceptions, setting realistic and personally relevant goals whether they are performance-based for healthy individuals or focus on modifying sedentary routines for those with chronic conditions to maximize perceived ease of change and, consequently, engagement and adherence (McAuley et al., 2003; Neupert et al., 2009; Michie et al. 2009; Hutchesson et al. 2015; Rebar et al., 2016; Schau et al. 2025).

Associations with eating- and activity-based social comparison

We identified several positive associations between ratings of eating behaviors and PA, respectively, as healthier than that of others and perceived ease of changing related eating and PA behaviors. Indeed, individuals who perceive their current behaviors as superior to those of their peers may develop a stronger sense of self-efficacy via mastery experiences (Bandura, 1997; Strecher et al., 1986), believing that, since they are already doing well, further improvements are more easily achieved. Furthermore, social comparison theory suggests that downward social comparison can boost self-esteem and confidence, further reinforcing the belief in one's capability to succeed (Spray et al., 2023; Strecher et al., 1986).

When asked about comparison frequency, associations with perceived ease of changing diverged between the two study populations. For the young adult sample, a higher comparison frequency correlated with increased perceived ease of eating more, which could be influenced by prevailing social norms and hedonic goals within their peer groups (Stok et al., 2016). In many social settings, particularly among young adults, eating larger portions or indulgent foods might be normalized or even encouraged within social gatherings, leading individuals to perceive such behaviors as less effortful or more socially acceptable (Sogari et al., 2018). For older adults, frequent social comparisons were associated with perceived ease of eating less and healthier, which suggests a different comparative framework. In this group, social comparison may involve observing peers or support networks effectively managing their conditions through dietary discipline. Such upward comparisons can provide examples of success and enhance self-efficacy for one's own health-promoting behaviors, making more restrictive or healthier eating goals seem more achievable (Bandura 1997; Buunk, B. P., & Ybema, J. F. 1997; Arigo et al. 2020). This highlights how health status and life stage shape the relevance and interpretation of social comparison information.

Associations with prior behavior change attempts

Prior behavior change attempts also had divergent associations in the two samples and for the two behaviors of interest. For the young adult sample, prior attempts to change eating behavior were associated with perceiving "undesired/unhealthy" behaviors to be easier to adopt. This aligns with self-regulation theories where repeated experiences of success or failure can shape future behavioral predictions (Ouellette & Wood, 1998); in this case, failed attempts may have induced a feeling of restriction being too difficult, or (excessive) consumption being much easier. In contrast, for the older adult sample, prior efforts might lead to a perceived ease of healthy eating, which suggests a different learning trajectory. For this group, previous engagement in behavior change may have provided valuable mastery experiences or increased perceived behavioral control. The seriousness of their health conditions might also enhance their intent to learn from past attempts, viewing any partial success as evidence of their capability to achieve healthier outcomes, thus increasing their self-efficacy for adaptive changes (Bandura, 1997; Kwasnicka et al., 2016; Sniehotta et al., 2014).

Interestingly, patterns were opposite for PA, where for the young adult sample, engagement in formal exercise programs was associated with greater perceived ease of changing PA, which may indicate that program participations has bolstered their self-efficacy and confidence. Indeed, structured programs can provide clear guidance, opportunities for skill mastery, and positive feedback, thereby enhancing perceived behavioral control and making specific activity goals seem more attainable (Ryan & Deci, 2017). For the older adult sample, however, participation in formal exercise programs was associated with perceiving becoming more physically active more difficult, suggesting that these programs might highlight the significant barriers and effort involved in managing PA alongside their health challenges (Bauman et al., 2012; Shields et al., 2012). Additionally, when asked to indicate the number of times that they had started a formal program that lasted more or less than 3 days, especially in the older adult sample, the more they indicated to have started such a program (lasting 3 or more days), it was associated with a higher perceived ease to change healthy aspects of eating behavior such as "eat less overall, eat no more than 50g added sugar/day", suggesting that repeated attempts may reflect persistence. Importantly, we only assessed whether participants had previously attempted to change their behavior without directly asking whether they were successful; the interpretation of the diverging patterns thus need to be confirmed in future research.

Associations with current behaviors

The observation that lower consumption of sweets and meat-based foods correlated with a perceived ease in reducing added sugar and overall meat intake, respectively, aligns with the principle that weaker existing habits are generally easier to alter or cease (Gardner et al., 2012) since they demand less cognitive effort to resist or reduce them further. Conversely, a higher frequency of consuming vegetable-based foods being associated with a perceived ease in meeting a 5-portions-a-day target highlights the reinforcement of positive habit formation. Consistent engagement in a desirable behavior builds self-efficacy via mastery experiences, making the continued or increased performance of that behavior seem less effortful and more attainable (Bandura, 1997; Ouellette & Wood, 1998).

The lack of association between moderate/high PA levels and the perceived ease of changing to "*engage in moderate-high PA for at least 300 min/week*" may suggest a ceiling effect for perceived ease among already active individuals with chronic conditions. While these participants demonstrate high self-efficacy for their current activity levels, the target of 300 minutes per week of moderate-to-vigorous PA is substantial, especially when managing ongoing health issues. For individuals already meeting a certain threshold, further increases may represent a disproportionate increase in perceived effort or a conflict with other life demands or symptom management, even if they are physically capable. Thus, “more” is not necessarily “easier”, as also the findings for highly active participants suggest. Their high PA level might already be close to their physiological or psychological limits, and any further increase could be perceived as demanding an unsustainable level of effort, risking pain, fatigue, or symptom exacerbation (Shields et al., 2012; Webb et al., 2022). Contrarily, “*being less active overall”* is less demanding and would reduce effort, in line with the concept of effort avoidance in decision-making around behavior, where less effortful options are often perceived as easier (Bustamante et al., 2024). Moreover, the finding that participants who spent more time sitting perceived it as difficult to “*spend less time sitting*” but easier to “*spend more time sitting*” strongly points to the power of habit strength and behavioral inertia (Gardner et al., 2024; Gardner et al., 2012). Sedentary behaviors, such as sitting, are often deeply ingrained in daily routines, frequently influenced by environmental cues (e.g., office work, home entertainment) and requiring little conscious effort to perform (Gardner et al., 2024; Gardner et al., 2012). Breaking such established habits or routines demands significant self-regulation and sustained effort, which may feel overwhelming.

Avenues for future research

Importantly, the present work was based on the assumptions of established health psychology theories that postulate that setting concrete goals is crucial for behavior change (HAPA- Schwarzer, 2008; GST - Latham & Locke, 1991). Another line of work has emphasized the effectiveness of “open goals” (or non-specific goals) in promoting PA and increasing positive psychological experiences, particularly for individuals who are not currently active (Clarke et al., 2025). Open goals, which involve exploring performance (e.g., "see how far you can walk in 6 minutes"), lead to higher interest in repeating the activity and higher motivation for future engagement, compared to SMART goals (specific, measurable, achievable, relevant, time-bound) for certain groups (Swann, 2012; Swann et al., 2020; Swann et al., 2025; Clarke et al., 2025). While the present research did not include these types of goals, replicating the findings based on the open goals account would be valuable.

Future research should also employ longitudinal designs to establish causality and track how perceptions of ease evolve with actual behavior change attempts and health outcomes. Moreover, larger sample sizes could allow for additional analysis i.e., regression analysis to test relationships between the perceived ease of change rating and different predictors also including several covariates. Investigating the underlying cognitive and affective processes (i.e., specific outcome expectancies, emotional responses to effort, habit strength) that drive these differential perceptions through qualitative methods or experimental designs would also be valuable. Furthermore, studies could explore the effectiveness of interventions explicitly tailored to these distinct perceptions of ease across different age and health groups, and examine how digital health technologies can facilitate personalized goal setting and feedback (Krukowski et al., 2024).

Study limitations

This study, while offering valuable insights, is subject to several limitations. A primary limitation is its cross-sectional nature, which precludes the testing of causal relations (Levin, 2006); the associations present in this work have to be interpreted with caution and directionality (ease of change influencing behavior or vice versa) remains to be tested.

The reliance on self-reported behaviors is also a limitation, given numerous accounts of overreporting of desirable and underreporting of undesirable behaviors (Paulhus, 1991; Stone, 2007). Nevertheless, the measures used are well validated and were selected based on their demonstrated validity. Still, the presented analyses should be replicated using objective measures, such as accelerometers for PA. Furthermore, the specific demographic and health profiles of the two study samples (young adults without chronic conditions vs. adults with chronic conditions) necessitate cautious interpretation of direct comparisons, as age and health status are inherently confounded variables influencing perceptions and behaviors. The demographic homogeneity of the samples with a predominance of White, highly educated individuals limits the generalizability of the findings and further research using representative samples is required. Additionally, in Study 1, it was not directly controlled for bots/scam responses, even though we do not expect the likelihood of such phenomena to be high as no incentives to take part in the study were offered (Kramer et al., 2014; Teitcher et al., 2015; Pozzar et al., 2020). In Study 2, bot control strategies were integrated by the data collection panel.

4.6. Conclusion

These findings highlight the critical need for tailored, person-cantered interventions. Health promotion strategies should move beyond a one-size-fits-all approach, recognizing that what constitutes an "easy" or "difficult" change varies significantly between individuals. For young adults, interventions might leverage their self-efficacy for specific, measurable gains and integrate social components. For individuals with chronic conditions, interventions should prioritize manageable, health-relevant changes, potentially focusing on reducing sedentary behavior or small, sustainable dietary shifts, while explicitly addressing potential barriers like pain and fatigue. Behavior change plans should ideally be co-created with individuals, allowing their perceptions of ease and difficulty to guide goal setting and foster greater self-regulatory processes (Michie et al., 2011; Hutchesson et al. 2015; Michie et al. 2009; Schau et al. 2025; Sniehotta, 2009).

5. Measurement reactivity in smartphone -based dietary ecological momentary assessment

Anila Allmeta [1,2], Alea Ruf [3], Andreas B. Neubauer[4,5], Laura M. König [1,3]

[1] Faculty of Life Sciences: Food, Nutrition and Health, University of Bayreuth, Germany
[2] Department Health and Prevention, Institute of Psychology, University of Greifswald, Germany
[3] Faculty of Psychology, University of Vienna, Austria
[4] DIPF-Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany
[5] Institute of Psychology, RWTH Aachen University, Germany

This chapter has been submitted in the Applied Psychology: Health and Well-Being Scientific Journal:

Allmeta, A., Ruf, A., Neubauer, A., & König, L. M. (2026). Measurement reactivity in smartphone-based dietary ecological momentary assessment. Retrieved from osf.io/preprints/psyarxiv/uw2gk_v1

Contribution:

AA: conceptualization, methodology, formal analysis, investigation, data curation, writing-original draft; AR: software, investigation, writing-review & editing; AN: conceptualization, writing-review & editing; LK: conceptualization, methodology, writing-review & editing, supervision.

5.1. Abstract

Objectives: Dietary Ecological Momentary Assessment (EMA) is popular for capturing real-time eating behaviors. Yet research participation effects such as measurement reactivity, i.e. changes to cognitions or behaviors due to being measured in a study, may impair data quality. This preregistered study investigated whether measurement reactivity occurred an eight-day event-based smartphone-based dietary assessment. It furthermore tested which aspects of eating behavior were most susceptible, and the moderating roles of active information processing (AIP), previous experience with diet-tracking apps, and the intention to change eating behavior were explored. Healthy behaviors were expected to decline and unhealthy behaviors to increase during the study period.

Methods & Results: Eighty-two participants (85.4% women; age $M = 22.88$, $SD = 3.91$) recorded all meals and snacks via photos and text descriptions. A total of 2,414 meals were recorded. Results indicated measurement reactivity for fruit and vegetable consumption (i.e., decline over the study period), but not for snacks. Behaviors considered easier to change were not more prone to measurement reactivity. Also, previous app experience and intentions to change behavior did not modulate the effects. However, AIP exerted a moderating effect on change in fruit and vegetable intake.

Conclusion: Thus, dietary EMA is not a neutral observational process but may serve as a light intervention that influences the intake of certain food groups. Researchers should account for measurement reactivity in study planning (e.g., by focusing on less reactive behaviors) or analytically (e.g., by discarding initial days of data or controlling for change across time).

5.2. Introduction

Smartphone-based Ecological Momentary Assessment (EMA; Shiffman et al., 2008) facilitates the repeated assessment of behavior and its determinants within real-time, real-life contexts, offering significant potential for the advancement of psychological theory (Scholz, 2019). Researchers frequently employ EMA to evaluate diverse psychological constructs such as motivational factors and positive or negative affective states alongside various health-related behaviors (Ottenstein & Werner, 2022; Perski et al., 2022; Wrzus & Neubauer, 2023). In contrast to traditional one-time self-report measures such as questionnaires that require participants to aggregate experiences across numerous occasions (Coughlin, 1990), EMA is less susceptible to recall bias. Consequently, the adoption of EMA over retrospective questionnaires may enhance the overall validity of psychological research (Shiffman et al., 2008).

However, substituting traditional self-report measures with EMA may not mitigate all potential biases. A systematic review by König, Allmeta, et al., (2022) suggests that measurement reactivity, i.e., changes in people's behavior, emotions, or cognitions due to being measured as part of a research project (French & Sutton, 2010), may be as prevalent in intensive longitudinal studies using digital devices as in traditional questionnaire-based research (Miles et al., 2020; Rodrigues et al., 2015). Such reactivity is problematic, as it can distort findings across observational, and interventional research. Ultimately, these distortions may lead to erroneous conclusions, resulting in misallocated resources and potential risks to public health, particularly if ineffective interventions are implemented as a result (French et al., 2021a). While numerous studies are available that test reactivity effects in physical activity, research on reactivity to the intensive longitudinal measurement of food intake remains scarce (König, Allmeta, et al., 2022), despite it being a prominent use case of EMA research (Perski et al., 2022). This present study aimed to close this gap by investigating (a) whether measurement reactivity occurred over an eight-day smartphone based dietary assessment, and whether measurement reactivity would differ depending on (b) the specific aspects of eating behavior or (c) person-level characteristics.

Reactivity to the intensive longitudinal measurement of health behaviors

Measurement reactivity effects have been detected in other health behaviors: Studies on the digital measurement of physical activity yielded, on average, a small to medium effect ($d = 0.3$; Cohen, 1992). One study also included the consumption of sugar-sweetened beverages, but no study was identified that investigated daily dietary intake (König, Allmeta, et al., 2022). EMA is also popular to assess dietary intake; more than 120 studies using dietary EMA have been published until 2020 and this figure has further risen since (König, Van Emmenis, et al., 2022). Also, Perski et al., (2022) identified 111 dietary behavior EMA studies, with the majority focusing on snacking (25.2%), followed by general food intake (20.7%). Measurement reactivity may be prevented through relatively simple changes to data collection and analysis procedures (e.g., discard the first days of collected data, see French et al., 2021, or to statistically control for the effect in the analytical model, see Bendtsen & McCambridge, 2021). These preventative measures, however, should be based on empirical evidence that is currently lacking.

Since the publication of the systematic review (König, Allmeta, et al., 2022), one study was published testing potential reactivity effects to dietary EMA. In this study, Whitton and colleagues aimed to identify demographic and psychosocial correlates of measurement error and reactivity bias, among adults who used a 4-day image-based mobile food record through which energy intake (EI) was measured. While the total sample showed a modest 3% daily decrease in EI, some participants who were labelled *Reactive Reporters* showed a substantial 17% decrease in EI per day of recording. Overall, Whitton et al. suggest that even "low-burden" image-based tracking methods are susceptible to significant reactivity (Whitton et al., 2023).

Moderators of measurement reactivity in dietary research

Eating is a complex behavior, consisting of many aspects such as whether an individual eats and what and how much is consumed per eating occasion (Bellisle, 2014; Ogden, 2011). In addition to focusing on overall consumption measured by EI, food-based dietary guidelines highlight the relevance of measuring the consumption of core food groups (Schäfer et al., 2025; U.S. Department of Agriculture & U.S. Department of Health and Human Services, 2020). Foods are furthermore commonly divided into rather “healthy” and rather “unhealthy”. This is also reflected in the broad range of dietary EMA tools used in research that may focus on a broad range of eating behavior indicators (see also König, Allmeta, et al., 2022, for an overview). In the present study, we therefore included a range of aspects of eating behavior to test for which measurement reactivity effects occur.

First, the distinction between healthy and unhealthy behaviors is relevant to the study of measurement reactivity, as stereotypically healthy and unhealthy behaviors are expected to follow different patterns. Healthy behaviors are more pronounced in the beginning of the study and drop off as the study continues, while unhealthy behaviors are reduced in the first days of the study and then increase (e.g., König, Allmeta, et al., 2022). We therefore included both healthy (i.e., eating five portions of fruit and vegetables) as well as unhealthy (i.e., skipping meals) aspects of eating behavior in our analysis to test for these different patterns. Second, prior research indicates that individuals perceive some aspects of eating behavior as easier to change (i.e., eating no more than 300g of meat/week or eating five portions of fruits and vegetables) than others (i.e., skipping meals or eating less overall) (Allmeta, Arigo, et al., 2026). We assume that measurement reactivity is stronger when the aspects of eating behavior are perceived as easier to change and weaker when they are perceived as difficult to change.

This assumption was made based on prior research in the context of physical activity, which produced more consistent evidence for reactivity to the measurement of steps – which can be relatively easily changed by walking some more – versus moderate-to-vigorous physical activity – which many people find challenging to include in daily routines (König, Allmeta et al., 2022). In the present study, the amount of fruits and vegetables and the amount of snacks consumed/day were considered as easy to change, whereas the total amount of food consumed/day and skipping meals were considered as aspects of eating behavior difficult to change.

Third, when recording dietary intake, different tools require different interactions, which may induce different levels of active information processing (AIP). AIP is defined as the active reflection on the information that the participants have to provide while recording their food intake using different recording features (i.e., photo-based, detailed written description) (Allmeta, Sutton, et al., 2026a). In other contexts, more AIP is related to better understanding of information and performance (e.g., Natter & Berry, 2005). Supported also by Control Theory (Carver & Scheier, 1982, 2002), it could be assumed that more AIP in dietary assessments could lead to more pronounced changes in eating behavior (see also Barta et al., 2012). We thus expected that participants will exhibit stronger measurement reactivity effects if the tracking process induces more AIP in them.

Fourth, as suggested in König, Allmeta, et al., (2022) and Whitton et al., (2023), characteristics of the participants may influence the extent to which measurement reactivity occurs. For instance, measurement reactivity might be more pronounced in individuals who behave in a comparably unhealthy way because they see a greater need for change (Barta et al., 2012). Similarly, intentions to change behavior in line with health recommendations might also induce stronger reactivity effects since the study might be a nudge to finally act in line with these intentions. Thus, we expect that measurement reactivity is stronger in participants who indicate stronger intentions to eat healthier during the study period.

Finally, familiarity with tracking health behaviors or experience in using behavioral tracking tools (i.e., diet tracking apps) might impact measurement reactivity as regular trackers may not experience increased awareness or novelty at the beginning of the study, thus be less prone to measurement reactivity (Arigo & König, 2024; König, Allmeta, et al., 2022). We expect that measurement reactivity is weaker in participants who have previous experience using diet tracking apps. The preregistered research questions and hypotheses (c.f. https://osf.io/vuxfd) are summarized in Table 1.[3]

Table 1
Research questions, hypotheses, and summary of main findings of the study.

	Research Question (RQ)	Hypothesis	Main Findings
1	Does measurement reactivity (MR) occur in an 8-day (EMA period) smartphone-based dietary assessment?	**1a**. The number of snacks consumed per day will increase over the EMA period	**1a.** The number of snacks consumed per day decreased over the EMA period
		1b. Intake of unhealthy food groups (e.g., sweet and savoury snacks) will increase over the study period	**1b.** The intake of unhealthy food groups decreased over the study period
		1c. Intake of healthy food groups (e.g., fruits, vegetables) will decrease over the study period	**1c.** The intake of healthy food groups decreased over the study period

[3] Additionally, we preregistered to exploratively compare the perceived complexity of the EMA study app, AIP, and perceived changes in eating-related cognitions, intentions, and behaviors due to the tracking process to data of a previous hypothetical online study (Allmeta et al., 2026). In brief, in this prior study, participants were randomly allocated to viewing one of the eight EMA smartphone dietary tracking apps in form of mock ups that differed in the type and number of tracking features they used. A subsample of this prior study ($N = 42$) received a mock-up of the app used in this present study (N = 82); we thus compared these two samples. Since the exploratory research question was part of the pre-registration but not directly relevant for the reactivity analysis, we report the measures, analysis, and results in the supplement.

2	Does the extent of the effect differ, depending on which aspects of eating behavior are studied?	**2a.** MR is stronger in healthy (amount of fruit and vegetables consumed/day) and unhealthy (amount of snacks in gram/day) aspects of eating behavior that are relatively easy to change	**2a.** MR was descriptively less strong for aspects of eating behavior considered easier to change
		2b. MR is less strong in healthy (amount of food intake/day) and unhealthy (number of skipped meals/day) aspects of eating behavior that are relatively difficult to change	**2b.** MR was descriptively stronger for amount of food intake; whereas no evidence of MR for the number of skipped meals was found
3	Is MR stronger in participants in whom the tracking process induced more AIP?	**3a**. If tracking induced more AIP in a participant, they will exhibit stronger MR effects	**3a.** AIP moderated MR strength in healthy food groups consumption: a statistically significant decrease only occurred in fruits and vegetables consumed, but not in the number and amount of snacks consumed
4	Is MR stronger in participants who indicate intention to change their eating behavior in the next 14 days?	**4a.** MR is stronger in participants who indicate that they intend to eat healthier in the next 14 days	**4a.** Intention to eat healthier did not moderate MR strength
5	Is MR less strong in participants who have previous experience in using diet tracking apps?	**5a.** MR is less strong in participants who have previous experience in using diet tracking apps	**5a.** Previous experience in using diet tracking app did not moderate MR strength

5.3. Methods

This study was preregistered on the Open Science Framework (OSF; https://osf.io/vuxfd), where also all study materials and data are available (https://osf.io/hg2dy). The study protocol was approved by the ethics committee of the University of Bayreuth, Germany. The study had two data collection waves, the first in Germany (sample 1) and the second in Austria (sample 2).

Sample

No formal power analysis was conducted, since the lack of prior research in this area and the complex statistical analyses did not allow to produce reliable estimates based on the literature. Instead, we considered the sample size from studies with similar designs focusing on other health behaviors such as physical activity, and to balance the requirement for sufficiently large samples and feasibility considerations, we set the target sample size to 80 participants.

In both data collection waves, adults aged 18 to 35 years could participate in the study. This age range was chosen due to a cover story that concealed the true purpose of the study by indicating that the aim of the study was to characterize young adults' food intake. Participants had to have access to an Android smartphone to use the study app (MovisensXS) which at the time was only available for Android operational systems. A limited number of Android smartphones could be loaned to participants for the duration of the study if they did not possess an Android smartphone and if they could travel to the respective campus to pick up and return the smartphone. In the first wave, the study app was offered in English and German; participants thus had to have good command of the English or German language to take part. In wave 2, the app was only offered in German and thus, good command of German was required. Sample 1 was recruited via online social media platforms (i.e., Twitter, Instagram), through flyers distributed at the University of Bayreuth and the University of Frankfurt, and by word of mouth. All participants who recorded at least one meal per day were eligible to enter a raffle and receive one of twenty 25€ Amazon vouchers. Sample 2 was recruited exclusively via the university's online participant pool. Participants were compensated with partial course credit.

Study design and procedure

An event-based EMA study was conducted over a period of eight consecutive days. Participants were asked to record all main meals and snacks they consumed, but not drinks, by providing a general description of the meal components and by taking photos of the meals and/or snacks in the study app (MovisensXS GmbH, Karlsruhe). Additionally, participants could report skipped meals and add any belated recordings. The inclusion of those options was shown to increase compliance in previous studies (König, Van Emmenis, et al., 2022; Ziesemer et al., 2020).

Participants in sample 1 were enrolled and received all instructions on the study and app set-up via email. They only met with a member of the study team if they required a loan smartphone to use the study app. Participants in sample 2 attended two individual, in-person sessions with trained study staff. In the first session, they were introduced to the study aim and procedure, received an Android smartphone if needed, installed the study app and completed the pre-study questionnaire. In the second session, which took place after the EMA recording period, they uninstalled the study app or returned the loaned smartphone if applicable. All participants provided written informed consent prior to starting the study.

The remainder of the study procedure was the same for both samples. First, participants completed a pre-study questionnaire and provided demographic and anthropometric information and answered questions regarding their dietary habits, intention to eat healthier, previous experience with diet tracking apps, and their usual meal and snack consumption frequency. Then, during the EMA recording period of eight consecutive days, they self-initiated the recordings of meals (breakfast; lunch; dinner; snack), skipped meals and belated recordings. At the end of each day, they received a summary of the meals and snacks they had recorded, together with a reminder to report any meals/snacks they might had skipped or missed to report. Additionally, they were asked to indicate whether they considered the current day typical. At the end of the EMA period, participants were asked to complete a post-study questionnaire that aimed to gather participants perceived ease of changing 13 eating aspects and perceived changes in eating behavior due to taking part in this study.

Measures

Demographic and anthropometric information

Participants were asked to indicate their age (in years), gender (male/ female/ other), and education level (i.e., some high school, to 8^{th} grade, some college, no degree, Bachelor´s degree, Master´s degree). Also, anthropometric information was collected, i.e., height (in centimetres, then transformed to meters) and weight (in kilograms), based on which BMI (kg/m^2) was calculated.

Healthy eating behaviors: intention and social comparison

Intention to eat healthier in the next 14 days was assessed with one item (*in the next 14 days, I intend to eat healthier*) using a 5-point scale (1 *strongly disagree* to 5 *strongly agree*). Social comparison regarding the healthiness of the participant´s eating behavior was assessed with one item (*relative to the average person of your gender, age, and professional/family situation, how healthy are your eating behaviors?)* using a 5-point scale (1 *much less healthy* to 5 *much healthier*). This item was based on those used in Arigo et al. (2023) and modified to address the context of the present study.

Experience with using diet tracking app

Participants' previous experience using diet tracking apps was assessed using a stage model approach (König et al., 2018), which categorizes participants into five stages based on the Precaution Adoption Process Modell (Weinstein et al., 1998; Weinstein & Sandman, 1992). Participants were asked whether they were thinking on using an app to track their food intake, and they could choose one of the five statements: 1-*I have never thought about using an app for that;* 2- *I have thought about using an app for that, but so far, I did not do it;* 3- *I have thought about using an app for that, but it is not necessary for me to do it;* 4- *I am currently using an app for that and intend to continue to use it;* 5-*I have used an app for that, but I do not use it anymore*. For the purpose of this study, the information was recoded into no previous experience (0; if options 1, 2, or 3 were selected) and previous experience with diet tracking apps (1; if options 4 or 5 were selected).

Frequency of meal and snack consumption

Participants were asked to indicate on how many days in a typical week they would consume breakfast, lunch, dinner and the number of snacks they would consume in a typical day.

EMA period: Assessment of food intake

Recording of meals and snacks. In the study app, participants first indicated the meal type (breakfast, lunch, dinner, snack). They then indicated the number of courses (e.g., main course and dessert = 2 courses) they were having. For each course, they were asked to take a photo of the plate; to add a brief description of the food they were about to consume; to indicate whether they shared their plate with anyone else (if "*yes*", indicate the proportion of the food that they consumed) and if there were any leftovers (if "*yes*", they were asked to take an additional photo of the leftovers).

Belated recordings. Participants had the option to record any meal or snack that they consumed during the day but did not record at the time they were consuming it. First, they clicked the "belated recording" option on the app's home screen, selected the meal/snack to report, and then indicated the approximate time they had consumed the meal/snack. Lastly, they provided a description of the meal/snack consumed. Since the recording did not occur at the consumption time, the option to take a photo was not possible.

Recording skipped meals. In the study app, participants had also the option to indicate whether they had skipped any meal or snack during the day. Firstly, they clicked on the "*skipped meals"* option on the app's home screen and then chose which meal or snack they had skipped.

Typicality of day. At the end of each EMA recording day, participants were asked to indicate whether the current day was a typical day on a 5-point Likert scale (1-*strongly disagree* to 5-*strongly agree*).

Post-study questionnaire

Active information processing (AIP). AIP was assessed with seven items using a 5-point Likert scale (1- *strongly disagree* to 5- *strongly agree*; Allmeta et al., 2026). For instance, the participants rated whether recording a meal/snack with the app required them to think about what they were about to eat or required them to know exactly how much they were about to eat. On average, AIP was 2.80 (SD = 0.64, N = 80); internal consistency was estimated as Cronbach´s α = .62.

Statistical Analysis

Preprocessing of EMA data to quantify food intake

For each recorded meal and snack, trained raters determined the foods and amounts consumed based on the photos taken at the beginning and end of the meal, together with the descriptions the participants provided. The trained raters followed the procedure outlined in König and Renner (2018). First, the individual foods were identified and listed. The weight of each food was estimated in grams based on a coding manual that contained photos of different portion sizes of commonly consumed foods. For packaged foods, information was retrieved directly from the product. All recorded foods were entered into the DGExpert software (Deutsche Gesellschaft für Ernährung e.V. (DGE), 2021). For each food, the corresponding item in the Bundeslebensmittelschlüssel (*BLS*, 2025) and BLS key were documented. The overall amount and kilocalories per meal were then exported from the software and entered into the dataset. In addition, the amounts per corresponding food group relevant to the analysis (fruits, vegetables, sweet and savoury snacks) were also recorded; the food groups were derived from German food-based dietary guidelines (Schäfer et al., 2025). Based on the meal-level data, we computed the following food intake variables for the analysis: *total amount of fruits and vegetables consumed per day in grams*[4]; *overall amount of food consumed per day in grams*; *amount of snacks consumed per day in grams* (sum of the amounts of sweet and savoury snacks in gram consumed on a given day). The *number of snacks per day* and *number skipped meals per day* were determined directly from the EMA recordings by counting the respective entries per day.

To determine interrater agreement, the data from eight randomly selected participants were coded by two raters. Interrater agreement was determined for the amount estimation (based on which the other outcomes were later calculated) at the ingredient level by using two-way random single-measure Intraclass-Correlations (ICC). An interrater agreement of 0.50 was considered fair, while an interrater agreement of 0.92 was considered excellent (Cicchetti, 1994; Walter & Rack, 2009; Tavakol & Dennick, 2011). Overall, there was an excellent interrater agreement of ICC = .98 for meal weight estimations in grams.

Inferential statistics

Multilevel regression analyses were conducted in RStudio (version 4.5.2.; Posit team, 2026) using the packages lme4 (Bates et al., 2025) and lmertest (Kuznetsova et al., 2020). Unless otherwise indicated, we followed the preregistered data analysis plan (c.f. https://osf.io/vuxfd).

Data was analysed using (generalized) multilevel modelling with days nested in participants; time was included as the primary predictor, using a continuous variable (coded from 0 on day 1 to 7 on day 8; Model A), to examine if there are linear changes in the

[4] Deviating from the preregistration, the amount of fruit and vegetables and the overall amount of food consumed per day were reported in grams instead of portions.

outcome over time. Models were calculated with and without random slopes; if random slopes models did not converge, only random intercept models were reported.

To test RQ1, we additionally computed models with time as a categorical variable (Model B), with day 8 (the last day of the EMA recording) serving as the reference category, resulting in seven regression coefficients. No random slopes were estimated in these models since treating time as a categorical predictor results in multiple dummy variables; estimating random slopes for each individual day would lead to an overly complex model structure that frequently suffers from non-convergence or over-parameterization, particularly if there is insufficient variance at the person level for each specific day-to-day comparison.

Separate models were computed for each of the outcomes of interest, all of which were aggregated per day. In all models, we added a dichotomous variable coding day of the week (coded 0 for Monday to Friday; coded 1 for Saturday and Sunday) as covariate to account for potential differences in the outcomes on weekdays vs. weekends. For all models, Type 1 error was set to .05.

RQ 1: Does measurement reactivity occur in an 8-day (EMA period) smartphone-based dietary assessment?

We computed models for the outcomes *number of snacks per day* (1a) and, *amount consumed of unhealthy* (i.e., sweet and savoury snacks; 1b) and *healthy* (i.e., fruits and vegetables; 1c) *food groups per day*. For the number of snacks (1a), a generalized multilevel model with Poisson distribution was estimated, given that the number of snacks is count data. For the amount of unhealthy and healthy food groups per day (1b and 1c), linear multilevel models were estimated.

Since in a few cases Models A with random slopes did not converge, especially for the count outcomes, and since conclusions drawn from models with random slope and without random slope did not differ, we deviate from our preregistration and only present the results of the random intercept models for consistency.

RQ 2: Does the extent of the effect differ, depending on which aspects of eating behavior are studied?

To answer this research question, the results of models 1b-A (amount of snacks consumed in grams) and 1c-A (amount of fruit and vegetables consumed in grams) were used. Additionally, we conducted the analysis also for the outcome variables *total amount of food consumed* and *number of skipped meals*, with time as a continuous predictor (c.f. Model A described above). Semi-partial correlations (*sr*) as standardized measures of effect size (cf. Arigo et al., 2021; Baga et al., 2024) were calculated for the associations between time and respective outcome and compared descriptively.

RQ 3: Is measurement reactivity stronger in participants in whom the tracking process induced more AIP?

A cross-level interaction between time (Level 1) and AIP (Level 2 – centred on the grand mean) was entered into the models A computed for RQ1 and reported for the three outcomes of interest.

RQ 4: Is measurement reactivity stronger in participants who indicate intention to change their eating behavior in the next 14 days?

A cross-level interaction between time (Level 1) and intention (Level 2 – centred on the grand mean) was entered into the models A computed for RQ1 and reported for the three outcomes of interest.

RQ 5: Is measurement reactivity less strong in participants who have previous experience in using diet tracking apps?

A cross-level interaction between time (Level 1) and previous experience (Level 2) was entered into the models A computed for RQ1 and reported for the three outcomes of interest.

5.4. Results

Sample description

In total, 82 participants completed the study. Forty participants completed the study in winter and spring 2023 (sample 1). Another 42 participants completed the study in the second wave of data collection between spring and summer 2024 (sample 2). For the total sample, mean age was 22.88 years (*SD* = 3.91), 85.4% of them were female, 54.9% indicated to be high school graduates and the average BMI was 22.8 (*SD* = 3.8). Socio-demographic characteristics of the full and both subsamples are presented in Table 2.

The majority of participants stated that they have not had previous experiences with using diet tracking apps (67.1%). On average, they rated the healthiness of their eating behaviors as about the same to somewhat healthier compared to their peers (43.9% - about the same, 35.4% - somewhat healthier). When asked if they intended to eat healthier in the next 14 days, 53.7% of participants neither agreed nor disagreed. Most participants stated to consume breakfast (54.5%), lunch (62.2%), and dinner (68.3%) almost every day in a typical week (on average five to six days/week). The number of snacks consumed per day mainly varied between 1 (35.4%), 2 (34.1%), and 3 (13.4%) snacks (*M* = 1.83, *SD* = 1.22).

Participants perceived *eat[ing] no more than 300-600 grams of meat per week* (*M* = 4.84, *SD* = 1.65) as the easiest to change aspect of eating behavior and *skip[ping] meals* (*M* = 2.69, *SD* = 1.44) as the most difficult aspect to change; see Table 3 in the supplementary material for the list of all 13 eating related aspects with the respective *M* and *SD* (higher value = perceived as easier to change). On average they perceived the study app as not too complex

(M = 2.61, SD = 0.66), perceived tracking induced moderate levels of AIP (M = 2.80, SD = 0.64), eating-related cognitions (M = 3.79, SD = 0.76), intention (M = 2.47, SD = 0.81), and behavior (M = 2.27, SD = 0.67).

Table 2
Sociodemographic characteristics of the sample

	Response Option	Full Sample (N = 82)	Sample 1 (N = 40)	Sample 2 (N = 42)
Age (M, SD)		22.88, 3.91	24.7, 4.8	21.14, 1.4
Gender	Female	85.4%	85%	85.7%
	Male	14.6%	15%	14.3%
	Other	-	-	-
Education Level*	1-Some high school, to 8th grade	-	-	-
	2-Some high school, no university entrance diploma	-	-	-
	3-High school graduate, university entrance diploma or the equivalent	54.9%	32.5%	76.2%
	4-Some college, no degree	13.4%	10%	16.7%
	5-Trade/technical/vocational training	1.2%	2.5%	-
	6-Bachelor’s degree	11%	15%	7.1%
	7-Master’s degree	17.1%	35%	-
	8-Doctoral degree	2.4%	5%	-
BMI (M, SD)		22.8, 3.8[1]	22.8, 3.8[1]	21.2, 2.5

Note. [1] Information missing for one participant.

Description of EMA data

Participants recorded a total of 2,414 meals with an average of 3.69 meals per day (SD = 1.55). On average, they consumed 686 grams of food per day (SD = 384; min = 570, max = 829), of which, on average, 244 grams (SD = 200) were fruits and vegetables. Participants recorded an average of 1.23 (SD = 1.21) snacks per day with a mean amount of 104 grams (SD = 137). Additionally, participants reported to have skipped an average of 0.25 (SD = 0.51) meals per day. The compliance with the end of day question regarding the typicality of the day (e.g., *was this a typical day for you)* was very low (11% answered every day; 39% did not answer any). The results for each of the eight days of EMA recording are presented in Table 5 in the supplementary materials.

RQ 1: Does measurement reactivity occur during the EMA period (8 consecutive days)?

In RQ 1a it was hypothesised that the number of snacks per day would increase over the EMA period. When considering time as a linear predictor (Model A), a statistically significant effect was observed $b = -.03$, $p = .027$ (see Table 6 in supplementary for all results). However, since the direction of the effect was negative, indicating a general decrease in snack consumption over the EMA period, the hypothesis was rejected. Model estimates indicate that participants consumed on average 1.19 snacks on day 1. For each day in the study, participants consumed, on average, 3.33% fewer snacks. When considering time as categorical predictor (Model B), there was no significant difference between day 8 and day 1, 2, 3, 4, 5, 6, and 7 in terms of daily snack count, $|b| < 0.27$, $p > .069$ (see Table 7 in supplementary for all results).

In RQ 1b it was hypothesized that the intake of unhealthy food groups, measured as the amount of sweet and savoury snacks in grams, would increase over the EMA period. When considering time as a linear predictor (Model A), a statistically significant effect was observed $b = -4.16$, $p = .042$. However, again, the change was not in the expected direction as the consumed amount of snacks in grams decreased over the EMA period. The hypothesis was thus rejected. Model estimates indicate that participants consumed on average 119.3 grams of snacks on day 1. When considering time as categorical predictor (Model B), there was no significant difference between the reference day (day 8) and the other days in terms of the amount in gram of snacks consumed, $|b| < 30.54$, $p > .125$. The model results are reported in Tables 6 and 7 in the supplementary material.

In RQ 1c, it was hypothesized that the intake of healthy food groups, measured as the amount of fruit and vegetables consumed in grams, would decrease over the EMA period. When considering time as a linear predictor (Model A), a statistically significant negative effect was observed (b = -8.12, p = .004) in line with our hypothesis. In brief, model estimates indicate that participants consumed on average 276 grams of fruits and vegetables on day 1. When considering time as categorical predictor (Model B), a statistically significant difference was observed between day 1 and the reference day (day 8), b = 77.85, p = .002. In brief, participants consumed on average 219 gram of fruits and vegetables on day 8 and 78 grams more on day 1. There was no statistically significant difference for days 2-7, $|b|$ < 50.98, p > .072 for all. The model results are reported in Tables 6 and 7 in the supplementary material.

RQ 2: Does the extent of the effect differ, depending on which aspects of eating behavior are considered?

To test RQ 2, we descriptively compared *sr* as a standardized measure of effect size. We predicted that measurement reactivity would be stronger for consuming fruits and vegetables (stereotypically healthy behavior) and for consuming snacks (snacking is considered a stereotypically unhealthy behavior) since these aspects of eating behavior were considered relatively easy to change in prior research. In contrast, we expected measurement reactivity to be weaker for the overall amount consumed (since eating less is a stereotypically healthy behavior) and number of skipped meals (with meal skipping being a stereotypically unhealthy behavior). For both the amount of fruit and vegetables consumed (t(570.57) = -2.91, p = .004, sr = 0.09) and the amount of snacks consumed (t(570.76) = -2.04, p = .042, sr = 0.07), which were considered relatively easy to change, srs were descriptively smaller than for the amount of food intake (t(80.96) = -4.07, p < .001, sr = 0.14), which was previously considered relatively difficult to change. Yet, for skipped meals, which was also considered rather difficult to change, we found no evidence for measurement reactivity effects.

RQ 3: Is measurement reactivity stronger in participants in whom the tracking process induced more active information processing (AIP)?

In RQ 3 it was hypothesised that if the tracking process induced more AIP, measurement reactivity effects in the outcomes of interest would also be stronger. When considering time as a linear predictor (c.f. model A conducted for RQ1), the number of snacks consumed as an outcome variable and AIP as a Level 2 moderator, there was no statistically significant cross-level interaction (b = -0.03, p = .252). Also, for the amount of snacks consumed in grams, there was no statistically significant cross-level interaction (b = -3.27, p = .317). However, for the amount of fruit and vegetables in grams, there was a statistically significant cross-level interaction (b = 9.08, p = .042). A simple slope analysis was conducted to follow up this finding. Results indicated that for participants with low AIP (1 *SD* below the sample mean), the statistically significant daily decline in intake remained (b = -13.81, p < .001), while for participants with high AIP, (1 *SD* above the sample mean) the fruit and vegetables intake decline was no longer statistically significant (b = -2.31, p = .56). All results are presented in Table 8 in the in the supplementary materials.

RQ 4: Is measurement reactivity stronger in participants who indicate intention to change their eating behavior in the next 14 days?

In RQ 4 it was predicted that measurement reactivity effects would be stronger in participants who indicated an intention to change their eating behaviors in the next 14 days. When considering time as a linear predictor (c.f. model A conducted for RQ1) and the intention to change eating behavior as a Level 2 moderator, there were no statistically significant cross-level interactions for the outcomes number of snacks (b = -.01, p = .797), amount of snack in grams (b = 0.85, p = .723) or amount of fruit and vegetables in grams (b =

4.44, p = .173). We, hence, found no support for the hypothesis that participants' intention to change eating behavior would moderate measurement reactivity effect, contrary to our hypothesis. All results are presented in Table 8 in the supplementary materials.

RQ 5: Is measurement reactivity less strong in participants who have previous experience in using diet tracking apps?

In RQ 5 it was hypothesised that measurement reactivity is weaker in participants who indicated to have had previous experience in using diet tracking apps. The variable previous experience in using diet tracking app was a dichotomous variable (0 = no previous experience, 1 = previous experience). When considering time as a linear predictor (c.f. model A conducted for RQ 1) and previous experience in using diet tracking apps as a Level 2 moderator, there were no statistically significant cross-level interactions for predicting number of snacks (b = 0.06, p = .063), amount of snack in grams (b = 6.68, p = .123) or the amount of fruit and vegetables in grams (b = -4.83, p = .420). Consequently, results provided no support for the hypothesis that previous experience in using diet tracking apps would be related to measurement reactivity. All results are presented in Table 8 in the supplementary materials.

5.5. Discussion

This research was designed to assess the presence of measurement reactivity in an eight-day smartphone-based dietary EMA. Overall, results regarding the presence of measurement reactivity were mixed. Evidence for measurement reactivity was found for the amount of fruits and vegetables consumed, which decreased during the study period. For the number of and amount in grams of snacks consumed, a statistically significant decline was found, but in the opposite direction as expected. This change is thus unlikely to be attributed to measurement reactivity. We also found little support for our hypothesis that measurement reactivity is stronger in eating behaviors that are easier to change. In addition, a moderating effect of AIP, but not of the intention to change eating behavior and the previous experience in using diet tracking apps on measurement reactivity were found for the amount of fruit and vegetables consumed. Participants with lower levels of AIP had a steeper decrease in comparison to participants with a higher level of AIP, where the amount of fruit and vegetable consumed was maintained over time. Measurement reactivity thus seems to be of relatively little concern in dietary EMA research, yet if specific aspects of food intake are studied, such as fruit and vegetable consumption, preventative measures may need to be explored.

Measurement reactivity in healthy and unhealthy behaviors

Prior studies suggested that measurement reactivity patterns depend on how behaviors are stereotypically perceived (König, Van Emmenis, et al., 2022). Accordingly, we expected an increase of unhealthy behaviors over the EMA period, based on the assumption that participants would initially suppress these behaviors due to wanting to adhere to consumption norms or social desirability, eventually returning to baseline (König, Van Emmenis, et al., 2022). At the same time, we assumed healthy behaviors to be more pronounced in the beginning and decrease as the study continues. Results only confirmed the latter assumption: fruit and vegetable consumption decreased over the eight-day EMA period. Participants reported approximately 78 grams more on day 1 compared to the final day of the study.

This "peak-and-decline" pattern is a hallmark of measurement reactivity for stereotypically healthy behaviors (French & Sutton, 2010). Participants may begin the study with high motivation to present as healthy eaters - a phenomenon known as social desirability bias in self-reporting often observed in intensive longitudinal designs (Adair, 1984; Hebert et al., 1995). Initially, the novelty of the digital tool and the awareness of being observed may prompt participants to align their behavior with internal or social norms regarding healthy eating (Barta et al., 2012). As the study progresses, a decline in the motivational salience of the monitoring process may lead to a return to baseline habits or a decrease in the diligence required to report healthy foods (Degroote et al., 2020; Fuller-Tyszkiewicz et al., 2013; Maugeri & Barchitta, 2019).

If measurement reactivity in dietary EMA indeed exists, the question arises as to how long it persists. In the context of physical activity, studies indicate that the effect may be short-lived since it typically only persists for the first one to two days of the study (Arigo & König, 2024). The present study suggests that the effect might be somewhat longer lasting for certain eating outcomes such as healthy food groups. Specifically, when we considered time as categorical (Model B), and the amount of fruits and vegetables consumed as outcome, we only found a significant difference between day 1 and day 8. Thus, as also recommended by French et al. (2021), researchers may want to consider discarding the initial days of collected data to obtain more accurate estimates, or to statistically control for potential reactivity effects. However, more studies are needed to replicate the findings both regarding the presence of measurement reactivity and regarding its development over time. This is especially important since the study may have been too short to accurately detect the endpoint in reactivity, given that snacking and total amount of food consumed only showed a linear decrease without a clear endpoint of the decline. The study duration may not have been long enough to distinguish between a temporary novelty effect of tracking and a sustained change in eating behavior. To better understand the duration of measurement reactivity, longitudinal studies extending beyond one week are needed to determine if the reduction in snack consumption persists or if participants eventually return to baseline habits.

From a theoretical perspective, an explanation for the results might come from the Control Theory (Carver & Scheier, 1982, 2002). The act of recording every snack likely decreased the prominence of "mindless" eating episodes. Rather than a temporary suppression, the intensive EMA protocol might have functioned as a continuous behavioral intervention. The daily reduction in snack count suggests that the self-monitoring inherent in behavioral EMA might have enhanced self-awareness. Contrary to other behaviors such as physical activity that are often monitored passively, there are no reliable passive sensing tools for food intake available that provide the same level of detail as (photo-based) dietary EMA; measurement reactivity effects thus may persist for much longer.

Importantly, the burden of an EMA design and under-reporting effects must be considered (Allmeta, Sutton, et al., 2026a; Ziesemer et al., 2020). Ziesemer et al. (2020), found that missing events are a significant challenge in dietary EMA, with a substantial proportion of all eating occasions potentially not being recorded by participants. They identified possible reasons for the omissions such as forgetting, situational constraints, technical issues and burden, suggesting that missing events are not random but rather influenced by the type of eating occasion and social context. Thus, the observed decreases in the consumption of different food groups might not reflect a true change in consumption, but rather under-reporting, where participants selectively omit recording snacks, which are often smaller and more frequent than main meals, to reduce the effort of documentation (Elliston et al., 2017).

Measurement reactivity across aspects of eating behavior

We furthermore examined whether the strength of measurement reactivity varied depending on the specific aspect of eating behavior under study. It was hypothesized that malleable behaviors - those perceived as easier to change, such as fruit and vegetables intake or snacking, would show stronger reactivity than more stable behaviors like total food intake or meal skipping patterns (Allmeta et al., 2026;). However, our findings partially challenged these assumptions, descriptively revealing that the measurement reactivity effect for these easier to change behaviors was not larger than the measurement reactivity of total food intake. Participants' perceptions on the ease of change of eating related aspects in this study were very similar to those in Allmeta et al., (2026), with meat, fruit and vegetable consumption and snacking seen as easier to change, whereas reducing total food intake and skipping meals were seen as more difficult to change. The unexpected findings thus are probably not due to diverging perceptions of ease to change.

Only the lack of significant reactivity for skipped meals aligns with our initial hypothesis and aligns with findings on habit formation (Gardner, 2015). Meal patterns are often deeply embedded in social and work routines, making them less susceptible to the short-term fluctuations caused by assessment reactivity. Because the triggers for skipping meals (e.g., lack of time, lack of appetite) are often external or biological, the mere presence of a smartphone tracker might be insufficient to alter the behavior. In contrast, total food volume, even if considered difficult to change in the long term, might be more susceptible to portion size adjustments in the short term.

Moderators of measurement reactivity

To further understand the mechanisms underlying measurement reactivity, we examined three potential moderators: AIP, intention to change eating behavior, and previous experience with diet tracking apps. Overall, our findings suggest that while AIP may play a role in reactivity to the measurement of healthy food consumption, intentions and past experience do not significantly alter the reactivity process.

Contrary to our prediction, AIP did not significantly moderate reactivity for snack consumption, but only for fruit and vegetable intake. Yet the direction of this effect was contrary to our hypothesis. For participants with low AIP, we observed a significant decline in daily fruit and vegetable intake, whereas for those with high AIP, this decline was mitigated and became non-significant. This suggests that AIP may serve as a protective factor against measurement fatigue rather than a catalyst for reactivity. In dietary EMA studies, the number of recorded meals tends to decline over time; prior research suggest that this may not be due to changes in actual consumption but rather due to the high burden of recording and declining motivation (Degroote et al., 2020; Ziesemer et al., 2020). High levels of AIP imply more conscious engagement with the recording, which may sustain the participant's commitment to the recording process, ensuring that healthy behaviors like eating vegetables are consistently captured throughout the study duration rather than dropping off as the novelty fades. This also aligns with findings by Natter and Berry (2005) suggesting that active processing enhances performance stability.

Contrary to our expectations, participants' baseline intention to change their eating behavior did not moderate the strength of measurement reactivity. This finding is surprising, as one would expect those already motivated to change to be more sensitive to the "intervention-like" effects of diet tracking. This lack of moderation may be interpreted through the Intention-Behavior Gap (Sheeran & Webb, 2016). It might be possible that the gap is too large for a relatively simple EMA to bridge; even if participants intend to eat healthier, the mere act of recording does not provide them with the self-regulatory skills (e.g., action planning) necessary to translate those intentions into measurable behavioral shifts during the short study period. Importantly, to reduce user burden, we employed a relatively low effort recording process (Allmeta et al., 2026). Future research should test whether more burdensome recording processes may induce more measurement reactivity, and whether this may be even more pronounced in individuals with stronger healthy eating intentions.

While intentions represent a distal motivation, the reactivity observed in EMA may be a more proximal, bottom-up reaction to the monitoring process itself, occurring regardless of whether the participant had previously planned to change. This suggests that measurement reactivity in eating behavior may be a semi-automatic response (Eisele et al., 2023) to being observed rather than a tool utilized only by those already seeking to self-improve.

Finally, we also explored whether previous experience with diet tracking apps would buffer against measurement reactivity. It was hypothesized that experienced trackers would be less prone to the novelty and increased awareness that usually triggers measurement reactivity (Arigo & König, 2024). However, the results showed no significant interaction for any outcome variable, suggesting that measurement reactivity in dietary assessment may be context-specific. Even if a participant has used a commercial diet app (e.g., MyFitnessPal) in the past, the specific requirements of EMA research (e.g., photo-based tracking as in the present study) may be sufficiently different to trigger a new wave of awareness. The novelty effect of a scientific study protocol might override the habituation that may have occurred with personal tracking tools; however, this hypothesis warrants testing in future research.

Study limitations

An important limitation to acknowledge is the sample composition, with a predominance of White, highly educated female students, which limits the generalizability of the findings. Further research in more heterogeneous samples is required. Yet, these sample characteristics are common in EMA research (Perski et al., 2022; Wrzus & Neubauer, 2023), thus the present analysis still provides valuable insights into potential reactivity effects in the literature. Additionally, with a sample size of $N = 82$ the present design was not sufficiently powered to detect small or medium sized cross-level interaction effects. These null findings should therefore be considered particularly carefully. Testing for small effects of person-level characteristics on measurement reactivity requires substantially larger sample sizes (Mathieu et al., 2012), though we caution to consider the practical utility of finding small effects of person-level characteristics on measurement reactivity.

Furthermore, the sample included a high percentage of participants who already perceived their diet as somewhat healthier than their peers and had low initial intentions to change, which may limit the generalizability of the findings to populations with clinical metabolic needs or those with strong motivations for weight loss. Indeed, a history of significant weight loss (>10 kg) was identified as the strongest predictor of measurement reactivity, with these individuals having 3.4 times higher odds of reactive reporting in a previous study (Whitton et al., 2023) and thus, this information would be important to include in future research.

A previous systematic review of dietary EMA showed large heterogeneity in assessment protocols (König et al., 2022). The present study only implemented photo-based recordings. Although among the most frequent, results may not necessarily generalize to other protocols, such as using food databases. While the present study was among the first to study reactivity to the smartphone-based assessment of diet, we encourage more research on this issue to provide a more complete picture of the issue.

5.6. Conclusion

Our findings demonstrate that the act of recording intake via an EMA app is not a neutral process; rather, it may serve as a light intervention that may influence the intake of fruits and vegetables and snacks, while leaving core meal structures intact. The finding that active information processing moderates the intake of healthy foods highlights a vital link between digital monitoring and cognitive engagement. Practically, these results suggest that researchers using EMA for purely observational purposes may want to account for potential reactivity effects, especially when fruit and vegetable consumption are of interest. Solutions may include focusing on less reactive behaviors, discarding the first few days of data (French et al., 2021), or by controlling for the effects statistically (Bendtsen & McCambridge, 2021). Yet, other biases to dietary EMA, such as an overall decline in recordings due to the high user burden (Ziesemer et al., 2020), may be even more important threat to data quality and thus require special attention.

Acknowledgments

The authors thank Hanna Voit, Ronja Harrer for supporting the second wave of data collection in Austria and Suad Saliju for supporting the pre-processing and the analysis of the data.

6. General discussion

The overarching objective of this dissertation was to provide a granular investigation into the presence and potential moderators of measurement reactivity within the context of ecological momentary assessment of dietary intake through digital tools such as smartphones. Following a systematic trajectory to address the specific research questions, firstly a systematic review and meta-analysis was conducted to establish the empirical foundation of the dissertation.

Within the broader context of the dissertation, the findings provide a psychological "roadmap" for predicting MR. The systematic review and meta-analysis in the first study established that reactivity occurs also in digital in-the-moment assessments, whereas study two showed that the tool's complexity dictates the active information processing. Instead study three, identified the specific aspects of behaviors most susceptible to being influenced by the assessment itself. By establishing a hierarchy of aspects of behaviors based on the perceived ease of change, we could theorize that MR is not a uniform "blanket effect" across all health domains. Instead, reactivity is likely "selective", occurring primarily in high perceived ease of change aspects of a behavior where the mere act of being prompted by a smartphone-based tool provides just enough of a cognitive cue to overcome a low barrier to change. This integration allows to move more toward a person-behavior-tool interaction model of reactivity.

The final study served as the empirical capstone of the dissertation closing the loop append by the previous studies. It suggested that dietary EMA is to a certain extent reactive, and that MR in EMA dietary intake, might be a product of the interaction between a specific tool, specific behavior (or specific aspects of a behavior), and specific individuals. An individual (*person*) with prior experience in using a complex eating related smartphone app *(tool*) to record a behavior they find easy to change (*behavior*), showed the highest levels of reactivity.

Together, these four studies advance the field by moving beyond the question of *if* MR exists, toward a better understanding of *how* and *for whom* digital monitoring alters (aspects of) health behaviors. The collective evidence from these four papers suggests that while ambulatory assessment is a powerful tool for health psychology, it is not a neutral observer. The act of digital monitoring functions as a minor behavioral intervention. In general, for researchers, this could implicate that EMA data should be interpreted with an understanding that the values recorded may represent a slightly optimized version of the participant's actual behavior in life. Instead for practitioners, this research opens the door to using MR as a low-cost, digital nudge to promote health behavior change.

6.1. Moderators of MR in Dietary Intake

The systematic review established that reactivity to digital in-the-moment assessment of health behaviors is a present, albeit subtle phenomenon; meta-analysis revealed small significant effects. Generally, it showed that MR was not uniform in all health behaviors; it appeared more pronounced in studies targeting PA or alcohol consumption. The characteristics of the assessment (i.e., frequency of prompts) showed an inconsistent relationship with reactivity, suggesting that the mere presence of a digital monitoring device might be as influential as the frequency of the questions themselves. Specifically, being monitored digitally led to slight improvements in health-promoting behaviors (e.g., increased PA) and slight decreases in health-risk behaviors (e.g., reduced alcohol intake). Even though EMA in dietary intake is very popular, no studies investigating reactivity to dietary EMA were identified, revealing an important literature gap that requires further research.

Beyond the knowledge on the presence of MR, several possible moderators that might amplify the effect were identified in the systematic review and tested in subsequent studies of this dissertation. Specifically, AIP, ease of change behavior, familiarity with diet tracking and intention to change behavior.

Active Information Processing

Active information processing, awareness of being measured, and social desirability were identified as further possible moderators of MR. In this dissertation a central focus is laid to AIP as a possible psychological mechanism underlying MR. Unlike awareness; participants' awareness of being measured; which is often considered as a moderator to MR (McCambridge, Witton, et al., 2014), AIP represents a more granular, deliberate cognitive appraisal of the focal behavior during the assessment act, triggered from the interaction with the assessment tool. This interaction increases awareness and reflection on the behavior, potentially inducing measurement reactivity (Barta et al., 2012). AIP is related to a better understanding of information and performance (Natter and Berry, 2005) thus the required interaction with the assessment tool (AIP) might moderate MR. If a discrepancy between the current and the desired behavior is perceived, self-regulatory processes are triggered to adjust the behavior (Carver and Scheier, 1982; Kanfer & Gaelick-Buys, 1991), moreover external cues such as self-monitoring devices influence MR as a result of self-monitoring (Nelson and Hayes, 1981). Our findings suggest that AIP serves as a catalyst for reactivity specifically within the domain of fruit and vegetable consumption, indicating that the cognitive depth involved in recording i.e., reflecting on the nutritional value or the portion size while recording, may drive behavioral adjustments. However, these effects appear behavior-specific, necessitating further empirical studies to determine if AIP functions similarly across broader dietary patterns or diverse health behaviors. Furthermore, because AIP implies sustained cognitive engagement, the habitual nature of intensive longitudinal designs like EMA may lead to habituation as participants gain familiarity with the protocol. Self-regulated learning (SRL) diary studies highlight that self-monitoring induces reactivity because it fosters reflective processing, but only when participants remain actively engaged with the content (Panadero et al., 2016). These strands support the dissertation's interaction model of reactivity, in which behavioral impact emerges from an interaction between the tool's affordances (burden, clarity, feedback), the user's cognitive engagement (AIP) and the perceived ability (perceived ease of change) with the targeted behavior. In the second study (Allmeta et al., 2026), the focus was shifted to the tool itself and the interaction of participants with the tool, confirming that dietary EMA tools are not interchangeable. Complex tools (that combine different number and types of tracking features) were associated as well as required higher levels of AIP. Interestingly, while complex tools provided more granular data, they were rated lower on usability. At the same time, the complexity of the EMA protocols was not perceived to differentially influence eating-related cognitions, intentions, and behaviors. Whereas, in last study, results suggested that AIP exerted a moderating effect on change in fruit and vegetable intake during the EMA period. These findings suggest a certain discrepancy between the perceived influence and the actual potential influence of AIP in complex dietary EMAs when tested in real-life. In smartphone-based dietary EMAs a certain involvement and interaction of the participant with the tool cannot be avoided as they have to at least initiate the recording process i.e., by taking a picture or describing the meal content (König et al., 2018; König, Van Emmenis, et al., 2021). Additionally, some assessment tools ask for added input even after participants have recorded their dietary intake for example "belated recordings" at the end of the day (Renner et al., 2018; Ziesemer et al., 2020) which again might increase AIP and potentially MR.

Consequently, future research needs to adopt a temporal perspective to investigate the dynamic fluctuations of AIP throughout the study period to understand how cognitive engagement evolves alongside measurement frequency. Additionally, it might be relevant to test in the context of dietary EMAs the suggestion by French et al., 2021 to restrict the access participants have on recorded data so to reduce MR.

Ease of Change

Measurement reactivity may be influenced by the specific indicator used for the target behavior, as shown with significant variation among PA studies. For example, research focusing on steps or activity counts; indicators known to be highly correlated; frequently reported reactive effects (Sartini et al., 2015). In contrast, studies assessing moderate to vigorous PA rarely observed such reactivity. This discrepancy suggests that steps and activity counts are more susceptible to change because they require less effort and are more easily integrated into daily routines (Cheval & Boisgontier, 2021). For instance, increasing step counts by walking from an earlier bus stop is more manageable than arranging a formal gym session during a busy day (König et al.,2022).

Although few studies have directly compared multiple indicators with results on MR presence being not conclusive (Motl & Dlugonski, 2011; Ho et al., 2013; Tinlin et al., 2018) these findings support the hypothesis that MR is more prevalent when a behavior is inherently easier to modify. These findings from the systematic review suggest that research should explicitly consider several aspects or indicators for the same behavior in order to differentiate in the required effort.

The third study of this dissertation (chapter-4) investigated whether indeed some aspects of behaviors are perceived to be easier to change than others. The primary findings suggested that individuals perceive some aspects of behaviors as easier or more difficult to change. Specifically, additive behaviors (e.g. eating more fruit and vegetables; spend more time sitting) were consistently rated as significantly easier to change than subtractive or restrictive behaviors (e.g., eating less overall; spending less time sitting). The results additionally suggested that the perceived ease of change of a behavior might be a dynamic construct influenced by past mastery experiences (previous attempts to change the behavior) and self-efficacy. Thus, individuals may be more likely to engage with monitoring and change-efforts when they feel the “entry barrier” to the behavior is low, so also easier to change. Nevertheless, when participants rated the perceived ease of change of 13 aspects of eating behavior in fourth study (chapter-4; Allmeta et al., 2026), the aspects of eating behavior considered easier to change (i.e., FV intake) were not more prone to MR. Instead, the lack of significant reactivity on eating aspect perceived as difficult to change (i.e., skipping meals) aligns with findings on habit formation (Gardner et al., 2015). Dietary patterns are frequently rooted in social and work routines, which makes them less suspectable to the temporary fluctuations often triggered by MR. Since the factors that lead for example to skipping meals, such as time constraints or biological cues like appetite, are typically external or physiological in nature (Lally & Gardner, 2011; Pendergast et al., 2016), the simple act of using a smartphone tracker may not be enough to pass the “entry barrier” and change these behaviors.

Familiarity with tracking

Familiarity with tracking health behaviors was considered another possible moderator to MR. Regular trackers might not experience increased awareness in the beginning of the study; the novelty of interacting with a tool that tracks their behavior is low; thus, they might be less prone to MR. When previous experience with nutrition apps was assessed in study 2 (Allmeta et al., 2026) it was not perceived as potentially having an impact on any of the investigated aspects of eating behavior (perceived changes on cognition, intention and behavior). Whereas, in the fourth study, previous app experience did not modulate the effect of MR. These findings suggest that the novelty effect of interacting with the tool is outlived quickly perhaps also because of the characteristics of ecological momentary assessment as a method itself. In EMA studies participants are usually asked to assess the same behavior several times per day for several consecutive days, thus the familiarization period is rapid, and participants' awareness shifts away from the measurement process as they habituate to it (Barta et al., 2012). These findings align with other health behaviors as well. For example, a study that isolated researcher observation among participants already using their own activity trackers found no change in daily steps when observation was introduced, suggesting that experienced trackers did not show additional reactivity once in a study (König et al.,2024). However, it is not often the case that prior experience or familiarity with the tracking device is reported in MR research, even more unusual that it is statistically controlled for potential effects. Thus, it is difficult to evaluate whether results in previous studies could have been influenced by a potential difference between experienced and novel users (König et al.,2024). As suggested by König et al., 2024, future studies, should enroll and directly compare experienced and novel monitor users in order to draw stronger conclusions (König et al., 2024) not only in PA but also on dietary research.

Intention to change behavior

Intention to change their health behavior in the beginning of the study; as a result of the discrepancies between their actual and desired behavior (Barta et al., 2012); was another possible moderator suggested in the systematic review and subsequently investigated in the studies of this dissertation. While not all participants intend to change their behavior upon enrollment, health-study participants are often more health-conscious than the general public due to self-selection bias (Haynes & Robinson, 2019; Nuzzo, 2021). This increased health consciousness may correlate with a greater predisposition toward following health recommendations. Consequently, researchers should assess and include “intention to change” as a covariate to account for intra-individual differences in MR. Nevertheless, in the fourth study of this dissertation, intentions to change behavior did not modulate the effect of MR, contradicting thus the assumption that motivated individuals are more sensitive to diet tracking, potentially highlighting the Intention-Behavior Gap (Sheeran & Webb, 2016). A simple EMA may lack self-regulatory components, such as action planning, needed to bridge this gap during brief study periods. Alternatively, reactivity may be a proximal, bottom-up response to the monitoring process itself rather than a distal, motivation-driven shift. This suggests that MR in eating behavior is a semi-automatic response to being observed (Eisele et al., 2023), occurring independently of a participant's conscious desire for self-improvement.

Beyond Dietary Assessment

A critical question for the field is whether these findings are unique to dietary assessment. Dietary intake is uniquely complex because it is frequent, socially embedded, and involves additive (eating more) and restrictive (eating less) components. While we observed reactivity in healthy behaviors (fruits/vegetables), it was not recognized in unhealthy behaviors (i.e., sugary snacks). This suggests that reactivity might be driven more by proactive goal pursuit (doing more of the healthy) than inhibitory control (doing less of the unhealthy), at least in short-term EMA burst (Harkin et al., 2016). Executive-function work supports a distinction between initiatory (eat more fruits and vegetables) and inhibitory (eat fewer unhealthy foods) components (Allom & Mullan, 2014; Bartholdy et al., 2016; Dohle et al., 2017; Byrne et al., 2021; Favieri et al., 2024). Better inhibitory control is linked to lower unhealthy snack or high-fat intake, but not necessarily to higher fruit/vegetable intake (Kakoschke et al., 2015; McGreen et al., 2023; Allom & Mullan, 2014; Limbers & Young, 2015). Conversely, updating/initiating functions and initiation skills uniquely predict higher fruit/vegetable consumption (Allom & Mullan, 2014; Limbers & Young, 2015). Trait self-control relates much more strongly to reducing unhealthy snacks than to increasing fruit intake, with effects mediated by habit (effortless inhibition) (Adriaanse et al., 2014). Lastly, evidence from self-regulation reviews indicates that specific mechanisms (e.g., self-monitoring of behavior vs. outcomes) map differently onto behavioral versus outcome change, and effects vary by behavioral domain (Hennessy et al., 2020; Spring et al., 2021). This selectivity supports the claim that discrepancy-reduction loops are not automatic for all monitored behaviors. Motivational research further suggests that prevention-oriented regulation (avoiding negatives) and promotion-oriented regulation (approaching positives) relate differently to eating patterns and intentions (Lalot et al., 2019; Pfattheicher & Sassenrath, 2014; Fuglestad et al., 2023). Together, these findings reinforce that EMA reactivity presence may preferentially enhance promotive, approach-oriented dietary behaviors (e.g., fruits/vegetables) rather than preventive, avoidance-oriented ones (e.g., snacks), especially over short digital monitoring bursts.

EMA as an assessment method is not diet-specific but rather broadly applicable in numerous health behaviors. For example, a large EMA review across 5 key health behaviors (diet, PA, alcohol, tobacco, sexual health), found high adherence and no systematic differences by behavior (Perski et al., 2022). Multiple EMA studies in smoking, alcohol, and e-cigarette use show little or no behavioral reactivity on primary risk behaviors, even with intensive, long-term protocols (Cajita et al., 2023; Rowan et al., 2007; McCarthy et al., 2015; Hufford et al., 2002; Hoeppner et al., 2024). This aligns with findings that EMA of movement-related behaviors shows modest, short-lived reactivity mainly in the first days (Maher et al., 2024).

The mechanisms identified AIP and Perceived Ease of Change are likely generalizable to other consumption behaviors like alcohol and tobacco, though the direction and magnitude may differ. Alcohol consumption similarly to unhealthy behaviors (i.e., sugary snacks), it is a restrictive behavior and, is often monitored in EMA with “event-contingent” prompts (logging a drink as it happens). Whereas smoking is a highly habituated, almost automatic behavior (Orbell & Verplanken, 2010; Marteau et al., 2012; Baker et al., 2025). Our findings on AIP would suggest that reactivity in smoking studies might be lower because the behavior is less “cognitively accessible” or more “resistant to ease-of-change” perceptions (Rowan et al., 2007; McCarthy et al., 2015; Kochhar et al., 2024). However, if an EMA tool requires the participant to count “puffs” (high complexity/high AIP), we would expect significant reactivity (Kochhar et al., 2024).

Across alcohol and tobacco, reviews emphasize that MR is typically modest but can emerge when behaviors are targeted for change, effortful to monitor, or closely tied to motivations and self-regulation (Shiffman, 2009; 2014; Hoeppner et al., 2024; Kalina et al., 2025; McCarthy et al., 2015). EMA often increases awareness of internal states (craving, affect, self-efficacy) and captures rapid within-day changes, consistent with an AIP-based account of digital reactivity (Shiffman, 2009; 2014; McCarthy et al., 2015).

In summary, while the findings of this dissertation are mainly diet-specific, the process of digital reactivity, driven by the interaction of cognitive load, behavioral ease of change, and AIP, provides a versatile framework for other ambulatory health behavior research, beyond dietary intake.

6.2. Implication for Theory

The findings of this dissertation offer a critical evaluation of the theoretical frameworks traditionally used to explain MR. Most notably, our results provide partial support for the Self-Regulation Model (SRL) (Ouellette & Wood, 1998). While the SRM suggests that any monitoring should trigger a discrepancy-reduction loop, our findings from the fourth study indicate that this loop is highly selective. The fact that reactivity was observed for fruit and vegetable intake but not for snack consumption suggests that theoretical models must incorporate the valence and perceived effort of the behavior. It appears that reactivity is more likely to align with “promotive” self-regulation (striving for a positive goal) rather than “preventive” self-regulation (avoiding a negative behavior) in the context of brief, digital monitoring.

Furthermore, our results challenge the Cognitive Load Theory (Ajzen, 1991) in its simplest form. While the second study confirmed that complex tools increase cognitive load, this load did not translate directly into a linear increase in MR. This suggests a "U-shaped" relationship: a minimum level of AIP is required to trigger a behavioral shift, but excessive complexity may lead to "disengagement" rather than "heightened awareness". Theoretically, this suggests an interaction model of reactivity which integrates the tool's affordances with the participant's cognitive engagement as a possibly more robust framework than models focusing solely on the dose or frequency of assessment.

The dissertation findings can be framed as a refinement of contemporary self-regulation and measurement-reactivity theories, integrating evidence from digital health and self-monitoring.

6.3. Implications on Research

For the research community, these findings serve as both a warning and a guide. The primary implication is that ambulatory data should not be treated as a purely "naturalistic" baseline. Small but consistent behavioral effects of self-monitoring in diet and lifestyle interventions support this concern (Teasdale et al., 2018; Cavero-Redondo et al., 2020; Kassavou et al., 2021; Spring et al., 2018). If digital monitoring alters the very behavior it seeks to measure, researchers must implement "reactivity-aware" designs. Specifically, we recommend: a "burn-in" period where data is collected but excluded from final behavioral analyses to allow for initial reactivity to subside. EMA compliance and behavior often stabilize over days, and many trials analyze only defined assessment windows (e.g., one-week bursts) to avoid early adaptation effects (Wrzus & Neubauer, 2023; Spring et al., 2018). This calls for more transparent reporting of design choices that affect EMA data quality and interpretation (May et al., 2018; Wrzus & Neubauer, 2023; Stinson et al., 2022). Also, researchers might consider treating AIP as a standard process variable, measured alongside the primary outcomes.

In summary, for researchers, the findings of this dissertation argue for reactivity-aware EMA designs that accommodate early reactivity (e.g., burn-in), routinely measure AIP, and report tool burden and interface choices.

6.4. Implications for Practice: Is Reactivity an „Issue“?

In clinical and practical settings, the interpretation of measurement reactivity remains a double-edged sword: is reactivity an „issue“? From a purely psychometric perspective, yes, it introduces bias and threatens internal validity. However, from a public health perspective, MR could be a "free" intervention. The practical implication for practitioners is that they should not necessarily view the "optimized" data from an app as a failure of measurement, but as a reflection of the patient's maximal capacity for change. Digital self-monitoring often increases fruits/vegetables and reduces high-risk behaviors over the intervention period, even when effects are modest (Teasdale et al., 2018; Spring et al., 2018; Melo et al., 2025). These improvements can be interpreted as revealing an individual's achievable performance under support, rather than a failure of measurement.

To gain a more definite answer on whether MR is a temporary "novelty effect" or a sustainable behavioral shift, follow-up studies should follow. Long-term mHealth trials show that self-monitoring engagement typically declines over months, even as some behavioral gains persist (Spring et al., 2018; Cavero-Redondo et al., 2020; Kassavou et al., 2021). To determine whether EMA reactivity is a short-lived novelty effect or a gateway to habit formation, future work should employ long-term longitudinal tracking with "fading" prompts where monitoring intensity gradually decreases, to test whether initial EMA-induced changes are maintained, attenuate, or generalize once external prompts are reduced.

6.5. Strength and Limitations

As EMA are widespread in many fields of behavioral research, as well as dietary intake (Kwasnicka et al., 2021; Wrzus et al., 2022; Battaglia et al., 2022; Perski et al., 2022), this work addresses a significant gap in the literature. While movement-related behaviors (PA/SB) have dominated the MR discourse, this dissertation provides the first systematic investigation into smartphone-based dietary reactivity, filling a critical gap in nutritional epidemiology and digital health.

A cornerstone of this dissertation is its commitment to open science and data sharing initiatives. By pre-registering all studies on established platforms i.e., Open Science Framework (OSF), the work mirrors the rigorous procedures reducing analytic flexibility and bias. Moreover, the dissertation utilizes a robust research chain: starting with a systematic review to establish the evidence base, moving to experimental mock-up designs to isolate tool-related factors, exploring psychological moderators through cross-cultural surveys, and culminating in a well-powered, real-world EMA empirical test.

This research moves beyond the generic Hawthorne effect by introducing AIP as a granular psychological mechanism of MR. By distinguishing AIP from mere awareness, the dissertation provides additional lens through which to view how the cognitive depth of an assessment tool (e.g., reflecting on nutritional value during entry) directly drives behavioral adjustment.

Additionally, the introduction of ease of change provides a psychological roadmap for predicting where MR is most likely to occur. Identifying that additive/promotive behaviors (e.g., eating more fruit) are perceived as easier to change and more susceptible to reactivity than subtractive/inhibitory behaviors (e.g., eating less snacks) is a significant contribution to self-regulation theory.

The work successfully bridges the gap between lab-controlled perception (Study 2 & 3) and real-world application (Study 4). It offers actionable parameters for researchers, such as the necessity for the selection of tool complexity based on the desired balance of data granularity versus participant burden.

While the dissertation included data from multiple countries (Germany, Austria, USA) and addressed both healthy young adults and older adults with chronic conditions, the empirical capstone (Study 4) utilized a relatively small sample size (n=82) compared to large-scale epidemiological surveys. This may limit the generalizability of the findings to more diverse socio-economic or clinical populations. Moreover, although the research addresses the "run-in" period of reactivity, most assessments were restricted to short-term bursts (e.g., eight days). It remains unclear whether reactivity effects "reset" in longitudinal studies with repeated bursts or if long-term habituation eventually overrides the initial reactive peak.

Additionally, in the second study, mock-ups were used to investigate the different constructs, and reactivity was assessed via perceptions rather than actual dietary change. While this was necessary to isolate tool complexity, the transition from "perceived influence" to "actual behavior change" involves complex environmental and biological barriers that hypothetical scenarios cannot fully capture. Nevertheless, in the last study many of the same constructs were tested in real-word settings.

Lastly, the rapid evolution of smartphone technology means that usability and AIP are moving targets. The distinction between manual entry and photo-logs may blur as AI-driven automated food recognition becomes standard, potentially altering the nature of cognitive load in ways not fully captured by current tool-based categorizations.

7. Conclusions

To conclude, this dissertation fills in an important gap in (dietary) EMA literature and invites a reflection on the relationship between the observer and the observed in the digital age. We have moved from asking *if* digital measurement is reactive to understanding *how* the architecture of our tools and the nature of our behaviors conspire to change our actions. The findings of this dissertation suggest that MR is not merely a methodological hurdle to be overcome, but an important characteristic of human interaction with digital measurement tools. By synthesizing the results across four distinct studies, we can conclude that the act of measuring health behavior potentially is, in itself, a psychological intervention that modulates at the very least the user's perceptions in relation to the behavior. As a takeaway these findings suggest that human health behavior is inherently "observational-sensitive". The act of recording one's life on a smartphone is not merely a technical task; it is a moment of self-reflection. This research shows that when a tool is demanding enough to trigger AIP, and when the behavior is perceived as "easy" enough to change, the smartphone acts as a mirror that does not just reflect reality but subtly modulates it.

Moreover, this work moves the field beyond the simplistic categorization of "Hawthorne effects" toward a more nuanced person-behavior-tool interaction of reactivity. We have shown that the magnitude of reactivity is a product of the interaction between the *tool's complexity* (AIP), the *behavior's plasticity* (Perceived Ease of Change), and the *user's individual characteristics* (Intentions to change behavior and previous experiences).

Additionally, this dissertation suggests that in the era of digital health, the boundary between assessment and intervention is increasingly porous. If a simple photo-log can shift healthy behaviors while a complex database might trigger restrictive self-regulation, then researchers might need to treat the choice of an EMA tool as a "dosage" decision. Moving forward, the field should consider to even more embrace the paradox of digital trace. We use these tools also to increase the ecological validity of the assessment and so to capture the “truth” about behavior, yet the tools themselves make that truth a moving target. Rather than seeing this as a flaw, researchers might consider it as a bridge. Perhaps the “optimized” version of the self that emerges during an EMA study is not a “false” data point, rather a map of what is possible. For the researcher, this requires a new level of statistical humility. The field should transition from trying to “eliminate” MR to accounting even more for it. This might involve implementing washout periods in the first days of data collection and using the insights gained here to calibrate the sensitivity of clinical trials. For the clinician, it offers a powerful new lever. However, further research is necessary to better understand the duration of these washout periods in dietary EMA.

This research underscores that in health psychology, the observer and the observed are inextricably linked; our digital "mirrors" do not just reflect our reality but potentially and just perhaps inevitably modulate it.

References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology*, *27*(3), 379–387. https://doi.org/10.1037/0278-6133.27.3.379

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, *69*(2), 334–345. https://doi.org/10.1037/0021-9010.69.2.334

Ahmadi, A., Noetel, M., Parker, P., Ryan, R. M., Ntoumanis, N., Reeve, J., Beauchamp, M., Dicke, T., Yeung, A., Ahmadi, M., Bartholomew, K., Chiu, T. K. F., Curran, T., Erturan, G., Flunger, B., Frederick, C., Froiland, J. M., Gonz_alez-Cutre, D., Haerens, L., … Lonsdale, C. (2023). A classification system for teachers' motivational behaviors recommended in self-determination theory interventions. Journal of Educational Psychology, 115(8), 1158–1176. https://doi.org/10.1037/edu0000783

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2),179–211. https://doi.org/10.1016/0749-5978(91)90020-t

Allmeta, A., Arigo, D., & König, L. M. (2026). Easy or difficult? Investigating perceived ease of changing eating and physical activity behaviors. *Applied Psychology: Health and Well-Being*, *18*(1), e70124. https://doi.org/https://doi.org/10.1111/aphw.70124

Allmeta, A., Sutton, S., & König, L. M. (2026). The same, only different: Smartphone-based dietary Ecological Momentary Assessment tools vary in complexity, usability and active information processing. *British Journal of Health Psychology*, *31*(1), e70057. https://doi.org/10.1111/bjhp.70057

Ambrosini, G. L., Hurworth, M., Giglia, R., Trapp, G., & Strauss, P. (2018). Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. *Nutrition Journal*, *17*(1), 5.https:// doi. org/ 10. 1186/ s1293 7-018-0315-4

Arigo, D., & König, L. M. (2024). Examining reactivity to the measurement of physical activity and sedentary behavior among women in midlife with elevated risk for cardiovascular disease. *Psychology & Health*, *39*(3), 319–335. https://doi.org/10.1080/08870446.2022.2055024

Arigo, D., Bercovitz, I., Satish, A., Lapitan, E., Folk, A. L., & Lobo, A. F. (2025). Personalized and adaptive physical activity-based social comparison opportunities for women with health risks: Insights from a real-time test of the identification/contrast model. Journal of Health Psychology, 30(13), 3927–3945. https://doi.org/10.1177/13591053251326629

Arigo, D., Brown, M. M., Pasko, K., & Suls, J. (2020). Social comparison features in physical activity promotion apps: Scoping Meta-review. Journal of Medical Internet Research, 22(3), e15642. https://doi.org/10.2196/15642

Arigo, D., Mogle, J. A., & Smyth, J. M. (2021). Relations between social comparisons and physical activity among women in midlife with elevated risk for cardiovascular disease: An ecological momentary assessment study. *Journal of Behavioral Medicine*, *44*(5), 579–590. https://doi.org/10.1007/s10865-021-00229-7

Arigo, D., Roberts, S. R., & Butryn, M. L. (2023). Social comparisons between group members during behavioural weight loss treatment: Comparison direction, scale, and associations with weight loss maintenance. Psychology & Health, 38(4), 429–444. https://doi.org/10.1080/08870446.2021.1967953

Arigo, D., Suls, J. M., & Smyth, J. M. (2014). Social comparisons and chronic illness: Research synthesis and clinical implications. Health Psychology Review, 8(2), 154–214. https://doi.org/10.1080/17437199.2011.634572

Baga, K., Salvatore, G. M., Bercovitz, I., Mogle, J. A., & Arigo, D. (2024). Daily social comparisons among women in midlife with elevated risk for cardiovascular disease: A within-person test of the identification/contrast model. *Applied Psychology: Health and Well-Being*, *16*(4), 1778–1800. https://doi.org/10.1111/aphw.12553

Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman. Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J. F., & Martin, B. W. (2012). Correlates of physical activity: Why are some people physically active and others not? Lancet, 380(9838), 258–271. https://doi.org/10.1016/S0140-6736(12)60735-1

Baker, T., Kaye, J., & Piasecki, T. (2025). Theory-based strategies to enhance interventions for smoking: Relevance to a digital intervention.. *Behaviour research and therapy, 193*, 104812. https://doi.org/10.1016/j.brat.2025.104812

Barta, W. D., Tennen, H., & Litt, M. D. (2012). Measurement reactivity in diary research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life: Chapter 6: Measurement reactivity in diary research*. Guildford Press.

Bartholdy, S., Dalton, B., O'Daly, O., Campbell, I., & Schmidt, U. (2016). A systematic review of the relationship between eating, weight and inhibitory control using the stop signal task. *Neuroscience & Biobehavioral Reviews, 64*, 35-62. https://doi.org/10.1016/j.neubiorev.2016.02.010

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., & Green, P. (2025). *Lme4: Linear Mixed-Effects Models using 'Eigen' and S4. In (Version 1.1-36.0)* [Computer software]. https://cran. r-project.org/web/packages/lme4/index.html

Battaglia, B., Lee, L., Jia, S. S., Partridge, S. R., & Allman-Farinelli, M. (2022). The use of Mobile-Based Ecological Momentary Assessment (MEMA) methodology to assess dietary intake, food consumption behaviours and context in young People: a Systematic review. *Healthcare*, *10*(7), 1329. https://doi.org/10.3390/healthcare10071329

Baumann, S., Groß, S., Voigt, L., Ullrich, A., Weymar, F., Schwaneberg, T., Dörr, M., Meyer, C., John, U., & Ulbricht, S. (2018). Pitfalls in accelerometer-based measurement of physical activity: The presence of reactivity in an adult population.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? Perspectives on Psychological Science, 2(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Béjar, L. M., Reyes, Ó. A., & García-Perea, M. D. (2018). Electronic 12-hour dietary recall (e-12HR): Comparison of a Mobile phone app for dietary intake assessment with a food frequency questionnaire and four dietary records. *JMIR mHealth and uHealth*, *6*(6), e10409. https:// doi. org/ 10. 2196/ 10409

Bell, B. M., Alam, R., Alshurafa, N., Thomaz, E., Mondol, A. S., de la Haye, K., Stankovic, J. A., Lach, J., & Spruijt-Metz, D. (2020). Automatic, wearable-based, in-field eating detection approaches for public health research: A scoping review. NPJ Digital Medicine, 3(1), 1–14. https://doi.org/10.1038/s41746-020-0246-2

Bellisle, F. (2014). Meals and snacking, diet quality and energy balance. *Physiology & Behavior*, *134*, 38–43. https://doi.org/10.1016/j.physbeh.2014.03.010

Bendtsen, M., & McCambridge, J. (2021). Causal models accounted for research participation effects when estimating effects in a behavioral intervention trial. Journal of Clinical Epidemiology, 136, 77–83. https://doi.org/10.1016/j.jclinepi.2021.03.008

Biel, J.-I.,Martin, N., Labbe, D., & Gatica-Perez, D. (2018). Bites 'n' bits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 1–33. https:// doi. org/ 10. 1145/ 3161161

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). Effect sizes based on means. In Introduction to metaanalysis (pp. 21–32). Wiley.

Boushey, C. J., Spoden, M., Zhu, F. M., Delp, E. J., & Kerr, D. A. (2017). New mobile methods for dietary assessment: Review of image-assisted and image-based dietary assessment methods. *Proceedings of the Nutrition Society*, *76*(3), 283–294. https://doi.org/10.1017/S0029665116002913

Bringolf-Isler, B. (2014). Reactivity to accelerometer measurement of children and adolescents. Medicine & Science in Sports & Exercise, 46(6), 1140–1146. https://doi.org/10.1249/MSS.0000000000000215

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J.-P.,Chastin, S., Chou, R., Dempsey, P. C., DiPietro, L., Ekelund, U., Firth, J., Friedenreich, C. M., Garcia, L.,

Bundeslebensmittelschlüssel (BLS). (2025). Max Rubner-Institut, Karlsruhe. https://blsdb.de/

Byrne, M., Tanofsky-Kraff, M., Lavender, J., Parker, M., Shank, L., Swanson, T., Ramirez, E., Lemay-Russell, S., Yang, S., Brady, S., Zenno, A., Chivukula, K., Kelly, N., & Yanovski, J. (2021). Bridging executive function and disinhibited eating among youth: A network analysis.. *The International journal of eating disorders*. https://doi.org/10.1002/eat.23476

Burke, L. E., Wang, J., & Sevick, M. A. (2011). Self-Monitoring in Weight Loss: A Systematic Review of the Literature. *Journal of the American Dietetic Association*, *111*(1), 92–102. https://doi.org/10.1016/j.jada.2010.10.008

Bustamante, L. A., Barch, D. M., Solis, J., Oshinowo, T., Grahek, I., Konova, A. B., Daw, N. D., & Cohen, J. D. (2024). Major depression symptom severity associations with willingness to exert effort and patch foraging strategy. Psychological Medicine, 54(15), 1–12. https://doi.org/10.1017/S0033291724002691

Buunk, B. P., & Ybema, J. F. (1997). Social comparison and occupational stress: The identification-contrast model. In B. P. Buunk & F. X. Gibbons (Eds.), Health, coping, and well-being: Perspectives from social comparison theory (pp. 359–388). Erlbaum.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111–135. https:// doi. org/ 10. 1037/ 0033-2909.92.1. 111

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111

Carver, C. S., & Scheier, M. F. (2002). Control Processes and Self-Organization as Complementary Principles Underlying Behavior. *Personality and Social Psychology Review*, *6*(4), 304–315. https://doi.org/10.1207/S15327957PSPR0604_05

Charness, N., & Boot, W. R. (2016). Handbook of self-regulation: Research, theory, and applications: Aging and self-regulation.Aging and self-regulation (Third ed.). The Guilford Press.

Cheval, B., & Boisgontier, M. P. (2021). The theory of effort minimization in physical activity. Exercise and Sport Sciences Reviews, 49(3), 168–178 https://doi.org/10.1249/JES.0000000000000252

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Clarke, M. M., Gargioli, D., Biddle, S. J., Rebar, A. L., Beauchamp, M. R., Jackman, P. C., Rosenbaum, S., Vella, S. A., & Swann, C. (2025). Longitudinal study of open goals in physical activity promotion: protocol for 'Open to Move.'. BMJ Open Sport & Exercise Medicine, 11(2), e002621. https://doi.org/10.1136/bmjsem-2025-002621

Clemes, S. A., & Deans, N. K. (2012). Presence and duration of reactivity to pedometers in adults. Medicine & Science in Sports & Exercise, 44(6), 1097–1101. https://doi.org/10.1249/MSS.0b013e318242a377

Clemes, S. A., & Parker, R. A. A. (2009). Increasing Our Understanding of Reactivity to Pedometers in Adults. *Medicine & Science in Sports & Exercise*, *41*(3), 674–680. https://doi.org/10.1249/MSS.0b013e31818cae32

Clemes, S. A., Matchett, N., & Wane, S. L. (2007). Reactivity: An issue for short-term pedometer studies? British Journal of Sports Medicine, 42(1), 68–70. https://doi.org/10.1136/bjsm.2007.038521

Clinton, S. K., Giovannucci, E. L., & Hursting, S. D. (2020). The World Cancer Research Fund/American Institute for Cancer Research third expert report on diet, nutrition, physical activity, and Cancer: Impact and future directions. Journal of Nutrition, 150(4), 663–671. https://doi.org/10.1093/jn/nxz268

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. Psychological Bulletin, 119(1), 51–69. https://doi.org/10.1037/0033-2909.119.1.51

Conner, M., & Sparks, P. (2005). The theory of planned behavior. In M. Conner & P. Norman (Eds.), Predicting health behaviour: Research and practice with social cognition models. Open University Press. 17580854, 2026, 1,

Conner, T. S., & Barrett, L. F. (2012). Trends in Ambulatory Self-Report: The Role of Momentary Experience in Psychosomatic Medicine. *Psychosomatic Medicine*, *74*(4), 327–337. https://doi.org/10.1097/PSY.0b013e3182546f18

Cook, P., Schmiege, S., McClean, M., Aagaard, L., & Kahook, M. (2012). Practical and Analytic Issues in the Electronic Assessment of Adherence. *Western Journal of Nursing Research*, *34*(5), 598–620. https://doi.org/10.1177/0193945911427153

Coughlin, S. S. (1990). Recall bias in epidemiologic studies. Journal of Clinical Epidemiology, 43(1), 87–91. https://doi.org/10.1016/0895-4356(90)90060-3

CQUniversity's Physical Activity Research Group. (2001). 10,000 Steps. Cquniversity's Physical Activity Research Group. https://www.10000steps.org.au/

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. Medicine and Science in Sports and Exercise, 35(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Craig, C. L., Tudor-Locke, C., Cragg, S., & Cameron, C. (2010). Process and treatment of pedometer data collection for youth: The Canadian physical activity levels among youth study. Medicine & Science in Sports & Exercise, 42(3), 430–435. https://doi.org/10.1249/MSS.0b013e3181b67544

Cumming, G., & Calin-Jageman, R. J. (2017). Introduction to the new statistics: Estimation, open science, and beyond. Routledge.

Dakin, S., & Arrowood, A. J. (1981). The social comparison of ability. Human Relations, 34(2), 89–109. https://doi.org/10.1177/001872678103400201

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. https:// doi. org/ 10. 2307/ 249008

Davis, R. E., & Loprinzi, P. D. (2016). Examination of accelerometer reactivity among a population sample of children, adolescents, and adults. Journal of Physical Activity and Health, 13(12), 1325–1332. https://doi.org/10.1123/jpah.2015-0703

Degroote, L., DeSmet, A., De Bourdeaudhuij, I., Van Dyck, D., & Crombez, G. (2020a). Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: A systematic review. International Journal of Behavioral Nutrition and Physical Activity, 17(1), 1–13. https://doi.org/10.1186/s12966-020-00932-9

Degroote, L., Hamerlinck, G., Poels, K., Maher, C., Crombez, G., D Bourdeaudhuij, I., Vandendriessche, A., Curtis, R. G., & DeSmet, A. (2020b). Low-cost consumer-based trackers to measure physical activity and sleep duration among adults in free-living conditions: Validation study. JMIR mHealth and uHealth, 8(5), e16674. https://doi.org/10.2196/16674

Deutsche Gesellschaft für Ernährung e.V. (DGE). (2021). *DGExpert (Gemeinschaftsverpflegung (GV) Netzwerkversion Version 2.0).* [Computer software]. DGE Medienservice. https://www.dge.de/ernaehrungsberatung/dgexpert/

Ding, H., Hu, G. L., Zheng, X. Y., Chen, Q., Threapleton, D. E., & Zhou, Z. H. (2015). The method quality of cross-over studies involved in Cochrane systematic reviews. PLoS ONE, 10(4), e0120519. https://doi.org/10.1371/journal.pone.0120519 doi.org/10.1123/jpah.2020-0331

Dorina, I., Mullan, B., Boyes, M., & McAlpine, T. (2024). What is behavioral complexity? Lay perceptions of characteristics of complex behavior. Behavioral Sciences, 14(8), 730. https://doi.org/10.3390/bs14080730

Dössegger, A., Ruch, N., Jimmy, G., Braun-Fahrländer, C., Mäder, U., Hänggi, J., Hofmann, H., Puder, J. J., Kriemler, S., &

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ, 315(7109), 629-634.https://doi.org/10.1136/bmj.315.7109.629

Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., Myin-Germeys, I., & Viechtbauer, W. (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. https://doi.org/10.1037/pas0001177

Elliston, K. G., Ferguson, S. G., & Schüz, B. (2017). Personal and situational predictors of everyday snacking: An application of temporal self-regulation theory. *British Journal of Health Psychology*, *22*(4), 854–871. https://doi.org/10.1111/bjhp.12259

Elliston, K. G., Ferguson, S. G., Schüz, N., & Schüz, B. (2017). Situational cues and momentary food environment predict everyday eating behavior in adults with overweight and obesity. *Health Psycholog y: Official Journal of the Division of Health Psycholog y, American Psychological Association*, *36*(4), 337–345. https:// doi. org/ 10. 1037/ hea00 00439

Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory Assessment—Monitoring Behavior in Daily Life Settings. *European Journal of Psychological Assessment*, *23*(4), 206–213. https://doi.org/10.1027/1015-5759.23.4.206

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https:// doi. org/ 10. 3758/ BRM. 41.4. 1149

Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7(2), 117–140. https://doi.org/10.1177/001872675400700202

Fogg BJ (2011). Tiny habits: The small changes that change everything. http://tinyhabits.com/

Fogg, B. J. (2009). A behavior model for persuasive design. In S. Chatterjee & P. Dev (Eds.), Proceedings of the 4th international conference on persuasive technology (pp. 1–7). ACM.

Foley, J. T., Beets, M. W., & Cardinal, B. J. (2011). Monitoring children’s physical activity with pedometers: Reactivity revisited. Journal of Exercise Science & Fitness, 9(2), 82–86. https://doi.org/10.1016/S1728-869X(12)60002-1

Foote, S. J., Wadsworth, D. D., Brock, S., Hastie, P., & Cooper, C. K. (2017). The effect of a wrist worn accelerometer on children’s in-school and out-of-school physical activity levels. Swedish Journal of Scientific Research, 33(3), 1–6.

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human-Computer Interaction*, *35*(6), 456–467. https://doi. org/ 10. 1080/ 10447 318. 2018. 1456150

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? British Journal of Health Psychology, 15(3), 453–468. https://doi.org/10.1348/135910710X492341

French, D. P., Miles, L. M., Elbourne, D., Farmer, A., Gulliford, M., Locock, L., Sutton, S., McCambridge, J., & the MERIT Collaborative Group. (2021). Reducing bias in trials from reactions to measurement: The MERIT study including developmental work and expert workshop. *Health Technology Assessment*, *25*(55), 1–72. https://doi.org/10.3310/hta25550

French, D. P., Miles, L. M., Elbourne, D., Farmer, A., Gulliford, M., Locock, L., Sutton, S., McCambridge, J., Bower, P. P., Clemes, D. S., Conner, P. M., Dundas, M. R., Elbourne, P. D., Eldridge, P. S., Farmer, P. A., French, P. D., Gamble, P. C., Gulliford, P. M., Kee, P. F., … Sutton, P. S. (2021). Reducing bias in trials due to reactions to measurement: Experts produced recommendations informed by evidence. *Journal of Clinical Epidemiology*, *139*, 130–139. https://doi.org/10.1016/j.jclinepi.2021.06.028

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*(4), 607–613. https://doi.org/10.1016/j.bodyim.2013.06.003

Gardner, B. (2015). A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. *Health Psychology Review*, *9*(3), 277–295. https://doi.org/10.1080/17437199.2013.876238

Gardner, B., Lally, P., & Rebar, A. L. (2024). Can-and should-automaticity be self-reported using a single item? A secondary analysis of 16 datasets. Applied Psychology. Health and Well-Being, 16(4), 2438–2457. https://doi.org/10.1111/aphw.12600

Gardner, B., Lally, P., & Wardle, J. (2012). Making health habitual: The psychology of 'habit-formation' and general practice. British Journal of General Practice: the Journal of the Royal College of General Practitioners,62(605), 664–666.https://doi.org/10.3399/bjgp12X659466

GBD 2019 Viewpoint Collaborators. (2020). Five insights from the global burden of disease study 2019. The Lancet, 396(10258), 1135–1159.https://doi.org/10.1016/S0140-6736(20)31404-5

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. Psychological Bulletin, 144(2), 177–197. https://doi.org/10.1037/bul0000127

Gichu, M., Jago, R., Katzmarzyk, P. T., … Willumsen, J. F. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. British Journal of Sports Medicine, 54(24), 1451–1462. https://doi.org/10.1136/bjsports-2020-102955

Godin, G., Sheeran, P., Conner, M., & Germain, M. (2008). Asking questions changes behavior: Mere measurement effects on frequency of blood donation. *Health Psychology*, *27*(2), 179–184. https://doi.org/10.1037/0278-6133.27.2.179

Goldschmidt, A. B., Wonderlich, S. A., Crosby, R. D., Engel, S. G., Lavender, J. M., Peterson, C. B., Crow, S. J., Cao, L., & Mitchell, J. E. (2014). Ecological Momentary Assessment of stressful events and negative affect in bulimia nervosa. *Journal of Consulting and Clinical Psycholog y*, *82*(1), 30–39. https:// doi. org/ 10. 1037/ a0034974

Götz, F.M., Gosling, S. D., & Rentfrow, P. (2022). Small effects: The indispensable foundation for a cumulative psychological science. Perspectives on Psychological Science, 17(1), 205–215. https://doi.org/10.1177/1745691620984483

Grimm, P. (2010). Social desirability bias. Wiley international encyclopedia of marketing.

Haegele, J. A., Zhu, X., & Bennett, H. J. (2020). Brief report: Reactivity to accelerometer measurement among adolescents with autism spectrum disorder. Journal of Autism and Developmental Disorders, 1–5. https://doi.org/10.1007/s10803-020-04742-7

Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Hamilton, W. K. (2018). MAJOR: Meta-Analysis Jamovi R (Version 1.2.0).

Haynes, A., & Robinson, E. (2019). Who are we testing? Self-selection bias in laboratory-based eating behaviour studies. Appetite, 141, 104330. https://doi.org/10.1016/j.appet.2019.104330

Hebert, J. R., Clemow, L., Pbert, L., Ockene, I. S., & Ockene, J. K. (1995). Social Desirability Bias in Dietary Self-Report May Compromise the Validity of Dietary Intake Measures. *International Journal of Epidemiology*, *24*(2), 389–398. https://doi.org/10.1093/ije/24.2.389

Heilala, V., Kelly, R., Saarela, M., Jääskelä, P., & Kärkkäinen, T. (2022). The Finnish version of the affinity for technology interaction (ATI) scale: Psychometric properties and an examination of gender differences. *International Journal of Human-Computer Interaction*, *39*(4), 874–892. https:// doi. org/ 10. 1080/ 10447 318. 2022. 2049142

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psycholog y*, *15*(Pt 1), 1–39. https:// doi. org/ 10. 1348/ 13591 0709X466063

Hennessy, E., Johnson, B., Acabchuk, R., McCloskey, K., & Stewart-James, J. (2020). Self-regulation mechanisms in health behavior change: a systematic meta-review of meta-analyses, 2006–2017. *Health Psychology Review, 14*, 42 - 6. https://doi.org/10.1080/17437199.2019.1679654

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. BMJ, 327(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hilgenkamp, T., Van Wijck, R., & Evenhuis, H. (2012). Measuring physical activity with pedometers in older adults with intellectual disability: Reactivity and number of days. Intellectual and Developmental Disabilities, 50(4), 343–351. https://doi.org/10.1352/1934-9556-50.4.343

Ho, V., Simmons, R. K., Ridgway, C. L., van Sluijs, E. M., Bamber, D. J., Goodyer, I. M., Dunn, V. J., Ekelund, U., & Corder, K. (2013). Is wearing a pedometer associated with higher physical activity among adolescents? Preventive Medicine, 56(5), 273–277. https://doi.org/10.1016/j.ypmed.2013.01.015

Höchsmann, C., & Martin, C. K. (2020). Review of the validity and feasibility of image-assisted methods for dietary assessment.

Höchsmann, C., Fearnbach, N., Dorling, J. L., Fazzino, T. L., Myers, C. A., Apolzan, J. W., & Martin, C. K. (2021). Preference, expected burden, and willingness to use digital and traditional methods to assess food and alcohol intake. *Nutrients*, *13*(10), 3340. https:// doi. org/ 10. 3390/ nu131 03340

Hutchesson, M. J., Rollo, M. E., Krukowski, R., Ells, L., Harvey, J., Morgan, P. J., Callister, R., Plotnikoff, R., & Collins, C. E. (2015). eHealth interventions for the prevention and treatment of overweight and obesity in adults: A systematic review with meta-analysis. Obesity Reviews: an Official Journal of the International Association for the Study of Obesity, 16(5), 376–392. https://doi.org/10.1111/obr.12268

International Journal of Obesity, 44(12), 2358–2371. https://doi.org/10.1038/s41366-020-00693-2

Jeong, M., Arigo, D., Jackson, D., Salvatore, G., & Mogle, J. A. (2025). Associations between comparative perceptions of memory functioning and broader social comparison tendencies among middle-aged and older adults. International Journal of Aging and Human Development, 914150251401562. Advance online publication. https://doi.org/10.1177/00914150251401562

Jung, J., Wellard-Cole, L., Cai, C., Koprinska, I., Yacef, K., Allman-Farinelli, M., & Kay, J. (2020). Foundations for systematic evaluation and benchmarking of a mobile food logger in a large-scale nutrition study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(2), 1–25. https:// doi. org/ 10. 1145/ 3397327

Kanfer, F. H., & Gaelick-Buys, L. (1991). Self-management methods. In F. H. Kanfer & A. P. Goldstein (Eds.), Helping people change: A textbook of methods (pp. 305–360). Pergamon Press.

Kaushal, N., & Rhodes, R. E. (2015). Exercise habit formation in new gym members: A longitudinal study. Journal of Behavioral Medicine, 38(4), 652–663. https://doi.org/10.1007/s10865-015-9640-7

Klenk, J., Peter, R. S., Rapp, K., Dallmeier, D., Rothenbacher, D., Denkinger, M., Büchele, G., & the ActiFE Study Group (2019). Lazy sundays: Role of day of the week and reactivity on objectively measured physical activity in older people. European Review of Aging and Physical Activity, 16(1), 1–4. https://doi.org/10.1186/s11556-019-0226-1

Kochhar, S., Scholten, H., Maciejewski, D., Pingel, M., & Luijten, M. (2024). A mixed-methods investigation of an ecological momentary assessment protocol for cigarette-smoking youth: Psychometric properties and participant experiences. *Drug and Alcohol Dependence Reports, 14*. https://doi.org/10.1016/j.dadr.2024.100314

König, L. M., & Renner, B. (2018). Colourful = healthy? Exploring meal colour variety and its relation to food consumption. *Food Quality and Preference*, *64*, 66–71. https://doi.org/10.1016/j.foodqual.2017.10.011

König, L. M., & Renner, B. (2019). Boosting healthy food choices by meal colour variety: Results from two experiments and a just-in-time Ecological Momentary Intervention. *BMC Public Health*, *19*(1), 975. https://doi.org/10.1186/s12889-019-7306-z

König, L. M., Allmeta, A., Christlein, N., van Emmenis, M., & Sutton, S. (2022). A systematic review and meta-analysis of studies of reactivity to digital in-the- moment measurement of health behaviour. *Health Psycholog y Review*, *16*(4), 551–575.https:// doi. org/ 10. 1080/ 17437 199. 2022. 2047096

König, L. M., Attig, C., Franke, T., & Renner, B. (2021). Barriers to and facilitators for using nutrition apps: Systematic review and conceptual framework. *JMIR mHealth and uHealth*, *9*(6), e20037. https:// doi. org/ 10. 2196/ 20037

König, L. M., Sproesser, G., Schupp, H. T., & Renner, B. (2018). Describing the process of adopting nutrition and fitness apps: Behavior stage model approach. JMIR mHealth and uHealth, 6(3), e55. https://doi.org/10.2196/mhealth.8261

König, L. M., van Emmenis, M., Nurmi, J., Kassavou, A., & Sutton, S. (2022). Characteristics of smartphone-based dietary assessment tools: A systematic review. *Health Psycholog y Review*, *16*(4), 526–550.https:// doi. org/ 10. 1080/ 17437 199. 2021.2016066

König, L. M., Pasko, K., Baga, K., Harsora, R., & Arigo, D. (2024). Isolating the role of researcher observation on reactivity to the measurement of physical activity. *Applied Psychology Health and Well-Being*, *17*(1), e12630. https://doi.org/10.1111/aphw.12630

Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, *11*(4), 415–425. https://doi.org/10.1037/1040-3590.11.4.415

Kramer, J., Rubin, A., Coster, W., Helmuth, E., Hermos, J., Rosenbloom, D., Moed, R., Dooley, M., Kao, Y., Liljenquist, K., Brief, D., Enggasser, J., Keane, T., Roy, M., & Lachowicz, M. (2014). Strategies to address participant misrepresentation for eligibility in web-based research. International Journal of Methods in Psychiatric Research, 23(1), 120–129. https://doi.org/10.1002/mpr.1415

Krukowski, R. A., Denton, A. H., & König, L. M. (2024). Impact of feedback generation and presentation on self-monitoring behaviors, dietary intake, physical activity, and weight: A systematic review and meta-analysis. International Journal of Behavioral Nutrition and Physical Activity, 21(1), 3. https://doi.org/10.1186/s12966-023-01555-6

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2020). *Package 'lmerTest' (Version 3.1-3)*. [Computer software]. http:// CRAN.R-project.org/package=lmerTest

Kwasnicka, D., Dombrowski, S. U., White, M., & Sniehotta, F. (2016). Theoretical explanations for maintenance of behaviour change: A systematic review of behaviour theories. Health Psychology Review, 10(3), 277–296. https://doi.org/10.1080/17437199.2016.1151372

Kwasnicka, D., Kale, D., Schneider, V., Keller, J., Asare, B. Y., Powell, D., Naughton, F., Hoor, G. a. T., Verboon, P., & Perski, O. (2021). Systematic review of ecological momentary assessment (EMA) studies of five public health-related behaviours: review protocol. *BMJ Open*, *11*(7), e046435. https://doi.org/10.1136/bmjopen-2020-046435

Labhart, F., Tarsetti, F., Bornet, O., Santani, D., Truong, J., Landolt, S., Gatica-Perez, D., & Kuntsche, E. (2020). Capturing drinking and nightlife behaviours and their social and physical context with a smartphone application – Investigation of users' experience and reactivity. Addiction Research & Theory, 28(1), 62–75. https://doi.org/10.1080/16066359.2019.1584292

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests And ANOVAs. *Frontiers in Psycholog y*, *4*, 863. https:// doi. org/ 10. 3389/ fpsyg. 2013.00863

Lally, P., & Gardner, B. (2011). Promoting habit formation. *Health Psychology Review*, *7*(sup1), S137–S158. https://doi.org/10.1080/17437199.2011.603640

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. Organizational Behavior and Human Decision Processes, 50(2), 212–247. https://doi.org/10.1016/0749-5978(91)90021-K

Leahey, T. M., Gokee LaRose, J., Fava, J. L., & Wing, R. R. (2011). Social influences are associated with BMI and weight loss intentions in young adults. Obesity (Silver Spring, Md.), 19(6), 1157–1162. https://doi.org/10.1038/oby.2010.301

Levin, K. A. (2006). Study design III: Cross-sectional studies. Evidence-Based Dentistry, 7(1), 24–25. https://doi.org/10.1038/sj.ebd.6400375

Lezhnina, O., & Kismihók, G. (2020). A multi-method psychometric assessment of the affinity for technology interaction (ATI) scale. *Computers in Human Behavior Reports*, *1*, 100004. https:// doi. org/ 10. 1016/j.chbr. 2020. 100004

Ling, F. C., Masters, R. S., & McManus, A. M. (2011). Rehearsal and pedometer reactivity in children. Journal of Clinical Psychology, 67(3), 261–266. https://doi.org/10.1002/jclp.20745

Ling, J., & King, K. M. (2015). Measuring physical activity of elementary school children with unsealed pedometers: Compliance, reliability, and reactivity. Journal of Nursing Measurement, 23(2), 271–286. https://doi.org/10.1891/1061-3749.23.2.271

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. American Psychologist, 57(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Maher, J. P., Arigo, D., Baga, K., Salvatore, G. M., Pasko, K., Hudgins, B. L., & König, L. M. (2024). Measurement reactivity in Ecological Momentary Assessment studies of movement-related behaviors. *Journal for the Measurement of Physical Behaviour*, *7*(1). https://doi.org/10.1123/jmpb.2023-0035

Marteau, T., Hollands, G., & Fletcher, P. (2012). Changing Human Behavior to Prevent Disease: The Importance of Targeting Automatic Processes. *Science, 337*, 1492 - 1495. https://doi.org/10.1126/science.1226918

Martin, C. K., Correa, J. B., Han, H., Allen, H. R., Rood, J. C., Champagne, C. M., Gunturk, B. K., & Bray, G. A. (2012). Validity of the Remote Food Photography Method (RFPM) for Estimating Energy and Nutrient Intake in Near Real-Time. *Obesity*, *20*(4), 891–899. https://doi.org/10.1038/oby.2011.344

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*(5), 951–966. https://doi.org/10.1037/a0028380

Maugeri, A., & Barchitta, M. (2019). A Systematic Review of Ecological Momentary Assessment of Diet: Implications and Perspectives for Nutritional Epidemiology. *Nutrients*, *11*(11), 2696. https://doi.org/10.3390/nu11112696

McAuley, E., Jerome, G. J., Marquez, D. X., Elavsky, S., & Blissmer, B. (2003). Exercise self-efficacy in older adults: Social, affective, and behavioral influences. Annals of Behavioral Medicine: a Publication of the Society of Behavioral Medicine, 25(1), 1–7. https://doi.org/10.1207/S15324796ABM2501_01

McCambridge, J., & Kypri, K. (2011). Can simply answering research questions change behaviour? Systematic review and meta analyses of brief alcohol intervention trials. PLoS ONE, 6(10), e23748. https://doi.org/10.1371/journal.pone.0023748

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

McCarthy, D. E., Minami, H., Yeh, V. M., & Bold, K. W. (2015). An experimental investigation of reactivity to ecological momentary assessment frequency among adults trying to quit smoking. Addiction, 110(10), 1549–1560. https:// doi.org/10.1111/add.12996

McCloskey, K., & Johnson, B. T. (2019). Habits, quick and easy: Perceived complexity moderates the associations of contextual stability and rewards with behavioral automaticity. Frontiers in Psychology, 10, 1556. https://doi.org/10.3389/fpsyg.2019.01556

McInnes, C., Carstairs, S. A., & Cecil, J. E. (2023). A qualitative study of young peoples' thoughts and attitudes to follow a more plant-based diet. Frontiers in Psychology, 14, 1196142. https://doi.org/10.3389/fpsyg.2023.1196142

Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life*. Guilford press.

Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: A meta-regression. *Health Psycholog y*, *28*(6), 690–701. https:// doi. org/ 10. 1037/ a0016136

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions. *Annals of Behavioral Medicine*, *46*(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterizing and designing behaviour change interventions. Implementation Science: IS, 6, 42. https://doi.org/10.1186/1748-5908-6-42

Michie, S., Wood, C. E., Johnston, M., Abraham, C., Francis, J. J., & Hardeman, W. (2015). Behaviour change techniques: The development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). *Health Technology Assessment*, *19*(99), 1–188. https://doi.org/10.3310/hta19990

Middelweerd, A., Mollee, J. S., Van Der Wal, C. N., Brug, J., & Te Velde, S. J. (2014). Apps to promote physical activity among adults: A review and content analysis. *International Journal of Behavioral Nutrition and Physical Activity*, *11*(1), 97. https://doi.org/10.1186/s12966-014-0097-9

Miles, L. M., Elbourne, D., Farmer, A., Gulliford, M., Locock, L., McCambridge, J., Sutton, S., & French, D. P. (2018). Bias due to Measurement Reactions in Trials to improve health (MERIT): Protocol for research to develop MRC guidance. Trials, 19(1), 1–8. https://doi.org/10.1186/s13063-018-3017-5

Miles, L. M., Rodrigues, A. M., Sniehotta, F. F., & French, D. P. (2020). Asking questions changes health-related behavior: An updated systematic review and meta-analysis. *Journal of Clinical Epidemiology*, *123*, 59–68. https://doi.org/10.1016/j.jclinepi.2020.03.014

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychological Science*, *7*(3), 221–237. https://doi.org/10.1177/1745691612441215

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., for the PRISMA G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097. https://doi.org/10.1371/journal. pmed.1000097

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Lisy, K., Qureshi, R., Mattis, P., & Mu, P. (2020). Chapter 7: Systematic reviews of etiology and risk. In E. Aromataris & Z. Munn (Eds.), JBI manual for evidence synthesis. JBI.

Morwitz, V. G., & Fitzsimons, G. J. (2004). The mere-measurement effect: Why does measuring intentions change actual behavior? Journal of Consumer Psychology, 14(1-2), 64–74. https://doi.org/10.1207/s15327663jcp1401&2_8

Morwitz, V. G., Johnson, E., & Schmittlein, D. (1993). Does Measuring Intent Change Behavior? *Journal of Consumer Research*, *20*(1), 46. https://doi.org/10.1086/209332

Motl, R. W., & Dlugonski, D. (2011). Increasing physical activity in multiple sclerosis using a behavioral intervention. Behavioral Medicine, 37(4), 125–131. https://doi.org/10.1080/08964289.2011.636769

Motl, R. W., McAuley, E., & Dlugonski, D. (2012). Reactivity in baseline accelerometer data from a physical activity behavioral intervention. Health Psychology, 31(2), 172–175. https://doi.org/10.1037/a0025965

Müller, A. M., Maher, C. A., Vandelanotte, C., Hingle, M., Middelweerd, A., Lopez, M. L., DeSmet, A., Short, C. E., Nathan, N., Hutchesson, M. J., Poppe, L., Woods, C. B., Williams, S. L., & Wark, P. A. (2018). Physical activity, sedentary behavior, and diet-related eHealth and mHealth research: Bibliometric analysis. Journal of Medical Internet Research, 20(4), e122. https://doi.org/10.2196/jmir.8954

Natter, H. M., & Berry, D. C. (2005). Effects of active information processing on the understanding of risk information. *Applied Cognitive Psycholog y*, *19*(1), 123–135. https:// doi. org/ 10. 1002/ acp. 1068 Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716. https:// doi. org/ 10. 1126/ scien ce. aac4716

Neupert, S. D., Lachman, M. E., & Whitbourne, S. B. (2009). Exercise self-efficacy and control beliefs: Effects on exercise behavior after an exercise intervention for older adults. Journal of Aging and Physical Activity, 17(1),1–16. https://doi.org/10.1123/japa.17.1.1

Nuzzo, J. (2021). Volunteer bias and female participation in exercise and sports science research. Quest, 73(1), 82–101. https://doi.org/10.1080/00336297.2021.1875248

Ogden, J. (2011). *The psychology of eating: From healthy to disordered behavior* (2. ed., [repr.]). Wiley-Blackwell.

Orbell, S., & Verplanken, B. (2010). The automatic component of habit in health behavior: habit as cue-contingent automaticity.. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association, 29 4*, 374-83. https://doi.org/10.1037/a0019596

Olander, E. K., Fletcher, H., Williams, S., Atkinson, L., Turner, A., & French, D. P. (2013). What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: A systematic review and meta-analysis. International Journal of Behavioral Nutrition and Physical Activity, 10, 29. https://doi.org/10.1186/1479-5868-10-29

Oscarsson, M., Carlbring, P., Andersson, G., & Rozental, A. (2020). A large-scale experiment on new year's resolutions: Approach-oriented goals are more successful than avoidance-oriented goals. PLoS ONE, 15(12), e0234097. https://doi.org/10.1371/journal.pone.0234097

Ottenstein, C., & Werner, L. (2022). Compliance in ambulatory assessment studies: Investigating study and sample characteristics as predictors. *Assessment*, *29*(8), 1765–1776. https:// doi. org/ 10. 1177/ 10731 911211032718

Ottenstein, C., Hasselhorn, K., & Lischetzke, T. (2024). Measurement reactivity in ambulatory assessment: Increase in emotional clarity over time independent of sampling frequency. *Behavior Research Methods*, *56*(6), 6150–6164. https://doi.org/10.3758/s13428-024-02346-y

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. Psychological Bulletin, 124(1), 54–74. https://doi.org/10.1037/0033-2909.124.1.54

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S.,…McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. BMJ, 372. https://doi.org/10.1136/bmj.n160

Paulhus, D. L. (1991). Measurement and Control of Response Bias. In Measures of personality and social psychological attitudes (pp. 17–59). Elsevier. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Pendergast, F. J., Livingstone, K. M., Worsley, A., & McNaughton, S. A. (2016). Correlates of meal skipping in young adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 125. https:// doi. org/ 10. 1186/ s12966-016-0451-1

Perski, O., Keller, J., Kale, D., Asare, B. Y.-A., Schneider, V., Powell, D., Naughton, F., ten Hoor, G., Verboon, P., & Kwasnicka, D. (2022). Understanding health behaviours in context: A systematic review and meta-analysis of Ecological Momentary Assessment studies of five key health behaviours. *Health Psycholog y Review*, *16*(4), 576–601. https:// doi. org/ 10. 1080/ 17437199. 2022. 2112258

Pew Research Center. (2020). About one-in-five Americans use a smart watch or fitness tracker. https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/

Pew Research Center. (2020). *Mobile Fact Sheet*. https:// www. pewre search. org/ inter net/ fact-sheet/mobile/

Posit team. (2026). *RStudio: Integrated Development Environment for R.* [Computer software]. Posit Software, PBC. http://www.posit.co/

Poulton, A., Pan, J., Bruns Jr, L. R., Sinnott, R. O., & Hester, R. (2019). A smartphone app to assess alcohol consumption behavior: Development, compliance, and reactivity. JMIR mHealth and uHealth, 7(3), e11157. https://doi.org/10.2196/

Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulsky, J. A., Hong, F., Gundersen, D. A., & Berry, D. L. (2020). Threats of bots and other bad actors to data quality following research participant recruitment through social media: Cross-sectional questionnaire. Journal of Medical Internet Research, 22(10), e23021. https://doi.org/10.2196/23021

Presseau, J., Tait, R. I., Johnston, D. W., Francis, J. J., & Sniehotta, F. F. (2013). Goal conflict and goal facilitation as predictors of daily accelerometer-assessed physical activity. *Health Psychology*, *32*(12), 1179–1187. https://doi.org/10.1037/a0029430

Prewitt, S. L., Hannon, J. C., & Brusseau, T. A. (2013). Children and pedometers: A study in reactivity and knowledge. International Journal of Exercise Science, 6(3), 230–235. Psychology, 34(1), 61–78. https://doi.org/10.1037/hea0000104

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(3), 253–270. https://doi.org/ 10.1080/15427609.2017.1340052

Rebar, A. L., Dimmock, J. A., Jackson, B., Rhodes, R. E., Kates, A., Starling, J., & Vandelanotte, C. (2016). A systematic review of the effects of non-conscious regulatory processes in physical activity. Health Psychology Review, 10(4), 395–407. https://doi.org/10.1080/17437199.2016.1183505

Reichenberger, J., Smyth, J. M., Kuppens, P., & Blechert, J. (2019). "I will fast … Tomorrow": Intentions to restrict eating and actual restriction in daily life and their person-level predictors. *Appetite*, *140*, 10–18. https:// doi. org/ 10. 1016/j. appet. 2019. 04. 019

Reichert, M., Giurgiu, M., Koch, E., Wieland, L. M., Lautenbach, S., Neubauer, A. B., von Haaren-Mack, B., Schilling, R., Timm, I., Notthoff, N., Marzi, I., Hill, H., Brüßler, S., Eckert, T., Fiedler, J., Burchartz, A., Anedda, B., Wunsch, K., Gerber, M.,…Liao, Y. (2020). Ambulatory assessment for physical activity research: State of the science, best practices and future directions. Psychology of Sport and Exercise, 101742. https://doi.org/10.1016/j.psychsport.2020.101742

Rodrigues, A. M., O'Brien, N., French, D. P., Glidewell, L., & Sniehotta, F. F. (2015). The question–behavior effect: Genuine effect or spurious phenomenon? A systematic review of randomized controlled trials with meta-analyses. *Health Psychology*, *34*(1), 61–78. https://doi.org/10.1037/hea0000104

Rosenstock, I. M., Strecher, V. J., & Becker, M. H. (1988). Social learning theory and the health belief model. Health Education Quarterly, 15(2), 175–183. https://doi.org/10.1177/109019818801500203

Rowan, P., Cofta-Woerpel, L., Mazas, C., Vidrine, J., Reitzel, L., Cinciripini, P., & Wetter, D. (2007). Evaluating reactivity to ecological momentary assessment during smoking cessation.. *Experimental and clinical psychopharmacology, 15 4*, 382-9. https://doi.org/10.1037/1064-1297.15.4.382

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluation. Developmental Psychology, 16(2), 105–115. https://doi.org/10.1037/0012-1649.16.2.105

Ruf, A., & König, L. M. (2025). What (not) to eat: Exploring weight-loss and dietary intentions in representative samples from Germany and Austria. Applied Psychology: Health and Well-Being. https://doi.org/10.31219/osf.io/bmr8t_v

Ryan, R. M., & Deci, E. L. (2017). Self-determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. Guilford Press.

Sallis, J. F., & Owen, N. (1999). Physical activity & behavioral medicine (Vol. 3). Behavioral medicine and health psychology series. Sage publications.

Sandberg, T., & Conner, M. (2009). A mere measurement effect for anticipated regret: Impacts on cervical screening attendance. *British Journal of Social Psychology*, *48*(2), 221–236. https://doi.org/10.1348/014466608X347001

Sartini, C., Wannamethee, S. G., Iliffe, S., Morris, R. W., Ash, S., Lennon, L., Whincup, P. H., & Jefferis, B. J. (2015). Diurnal patterns of objectively measured physical activity and sedentary behaviour in older men. BMC Public Health, 15(1), 1–13. https://doi.org/10.1186/s12889-015-1976-y

Sasaki, J. E., Da Silva, K. S., Gonçalves Galdino Da Costa, B., & John, D. (2016). Measurement of Physical Activity Using Accelerometers. In *Computer-Assisted and Web-Based Innovations in Psychology, Special Education, and Health* (pp. 33–60). Elsevier. https://doi.org/10.1016/B978-0-12-802075-3.00002-4

Sawyer, A. T., & McManus, K. (2021). Understanding patient experiences in a motivational interviewing intervention to improve whole-person lifestyle among individuals with hypertension or type 2 diabetes: A qualitative focus group study. International Journal of Qualitative Studies on Health and Well-Being, 16(1), 1978373. https://doi.org/10.1080/17482631.2021.1978373

Scandinavian Journal of Medicine & Science in Sports, 28(3), 1056–1063. https://doi.org/10.1111/sms.12977

Schäfer, A. C., Boeing, H., Gazan, R., Conrad, J., Gedrich, K., Breidenassel, C., Hauner, H., Kroke, A., Linseisen, J., Lorkowski, S., Nöthlings, U., Richter, M., Schwingshackl, L., Vieux, F., & Watzl, B. (2025). A methodological framework for deriving the German food-based dietary guidelines 2024: Food groups, nutrient goals, and objective functions. *PLOS ONE*, *20*(3), e0313347. https://doi.org/10.1371/journal.pone.0313347

Schau, N., Betz, C., Weber, A., Bader, A., König, L. M., Bartelmeß, T., Al Masri, M., Hofmann, R., & Haderer, M. (2025). HAPpEN: A pragmatic general practitioner-centered obesity management trial in rural Germany integrating multimodal lifestyle interventions. https://doi.org/10.21203/rs.3.rs-6540216/v1

Scholz, U. (2019). It's Time to Think about Time in Health Psychology. *Applied Psychology: Health and Well-Being*, *11*(2), 173–186. https://doi.org/10.1111/aphw.12156

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(6), 103. https:// doi. org/ 10. 9781/ijimai. 2017. 09. 001

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. Applied Psychology, 57(1), 1–29. https://doi.org/10.1111/j.1464-0597.2007.00325.x

Scott, J. J., Morgan, P. J., Plotnikoff, R. C., Trost, S. G., & Lubans, D. R. (2014). Adolescent pedometer protocols: Examining reactivity, tampering and participants' perceptions. Journal of Sports Sciences, 32(2), 183–190. https://doi.org/10. 1080/02640414.2013.815361

Seifert, A., Schlomann, A., Rietz, C., & Schelling, H. R. (2017). The use of mobile devices for physical activity tracking in older adults' everyday life. *DIGITAL HEALTH*, *3*, 2055207617740088. https:// doi. org/ 10.1177/ 20552 07617 740088

Sharp, D. B., & Allman-Farinelli, M. (2014). Feasibility and validity of mobile phones to assess dietary intake. *Nutrition*, *30*(11– 12), 1257–1266. https:// doi. org/ 10. 1016/j. nut. 2014. 02. 020

Sheeran, P., & Webb, T. L. (2016). The Intention–Behavior Gap. *Social and Personality Psychology Compass*, *10*(9), 503–518. https://doi.org/10.1111/spc3.12265

Shields, N., Synnot, A. J., & Barr, M. (2012). Perceived barriers and facilitators to physical activity for children with disability: A systematic review. British Journal of Sports Medicine, 46(14), 989–997. https://doi.org/10.1136/bjsports-2011-090236

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Smyth, J. M., Jones, D. R., Wen, C. K. F., Materia, F. T., Schneider, S., & Stone, A. (2021). Influence of Ecological Momentary Assessment study design features on reported willingness to participate and perceptions of potential research studies: An experimental study. *BMJ Open*, *11*(7), e049154. https:// doi. org/ 10. 1136/ bmjop en-2021-049154

Sniehotta, F. F. (2009). Towards a theory of intentional behaviour change: Plans, planning, and self-regulation. British Journal of Health Psychology, 14(Pt 2), 261–273. https://doi.org/10.1348/135910708X389042

Sniehotta, F. F., Presseau, J., & Araújo-Soares, V. (2014). Time to retire the theory of planned behaviour. Health Psychology Review, 8(1), 1–7. https://doi.org/10.1080/17437199.2013.869710

Sogari, G., Velez-Argumedo, C., G_omez, M. I., & Mora, C. (2018). College students and eating habits: A study using an ecological model for healthy behavior. Nutrients, 10(12). https://doi.org/10.3390/nu10121823

Spangenberg, E. R., Kareklas, I., Devezer, B., & Sprott, D. E. (2016). A meta-analytic synthesis of the question–behavior effect. *Journal of Consumer Psychology*, *26*(3), 441–458. https://doi.org/10.1016/j.jcps.2015.12.004

Spray, C. M., Franco, E., & Barnes, J. S. (2023). Social comparison in physical education: Impact of frames of reference, motives to compare, and self-determined motivation. Asian Journal of Sport and Exercise Psychology, 3(2), 65–73. https://doi.org/10.1016/j.ajsep.2023.06.001

Spring, B., Champion, K., Acabchuk, R., & Hennessy, E. (2020). Self-regulatory behaviour change techniques in interventions to promote healthy eating, physical activity, or weight loss: a meta-review. *Health Psychology Review, 15*, 508 - 539. https://doi.org/10.1080/17437199.2020.1721310

Sprott, D. E., Spangenberg, E. R., Block, L. G., Fitzsimons, G. J., Morwitz, V. G., & Williams, P. (2006). The question–behavior effect: What we know and where we go from here. *Social Influence*, *1*(2), 128–137. https://doi.org/10.1080/15534510600685409

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), Publication bias in meta-analysis: Prevention, assessment and adjustments (pp. 99–110). Wiley.

Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T.,…Higgins, J. P. T. (2019). Rob 2: A revised tool for assessing risk of bias in randomised

Stok, F. M., de Vet, E., de Ridder, D. T. D., & de Wit, J. B. F. (2016). The potential of peer social norms to shape food intake in adolescents and young adults: A systematic review of effects and moderators. Health Psychology Review, 10(3), 326–340. https://doi.org/10.1080/17437199.2016.1155161

Stone, A. A. (2007). The science of real-time data capture: Self-reports in health research. Oxford University Press.

Strecher, V. J., DeVellis, B. M., Becker, M. H., & Rosenstock, I. M. (1986). The role of self-efficacy in achieving health behavior change. Health Education Quarterly, 13(1), 73–92. https://doi.org/10.1177/109019818601300108

Sutton, S., Kinmonth, A.-L., Hardeman, W., Hughes, D., Boase, S., Prevost, A. T., Kellar, I., Graffy, J., Griffin, S., & Farmer, A. (2014). Does Electronic Monitoring Influence Adherence to Medication? Randomized Controlled Trial of Measurement Reactivity. *Annals of Behavioral Medicine*, *48*(3), 293–299. https://doi.org/10.1007/s12160-014-9595-x

Swann, C., Goddard, S. G., Schweickle, M. J., Hawkins, R. M., Williamson, O., Gargioli, D., Clarke, M. M., Jackman, P. C., & Vella, S. A. (2025). Defining open goals for the promotion of health behaviours: A critical conceptual review. Health Psychology Review, 19(2), 344–367. https://doi.org/10.1080/17437199.2025.2467695

Swann, C., Schweickle, M. J., Peoples, G. E., Goddard, S. G., Stevens, C., & Vella, S. A. (2020). The potential benefits of nonspecific goals in physical activity promotion: Comparing open, do-your-best, and as-well-aspossible goals in a walking task. Journal of Applied Sport Psychology, 34(2), 384–408. https://doi.org/10.1080/10413200.2020.1815100

Swann, W. B. (2012). Self-verification theory (pp. 23–42). SAGE publications ltd eBooks. https://doi.org/10.4135/9781446249222.n27

Szinay, D., Jones, A., Chadborn, T., Brown, J., & Naughton, F. (2020). Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review. *Journal of Medical Internet Research*, *22*(5), e17572. https://doi.org/10.2196/17572

Szymczak, H., Keller, L., Debbeler, L. J., Kollmann, J., Lages, N. C., Sproesser, G., Gollwitzer, P. M., Schupp, H. T., & Renner, B. (2020). “I'm eating healthy now”: The relationship between perceived behavior change and diet. *Food Quality and Preference*, *89*, 104142. https:// doi. org/ 10. 1016/j. foodq ual. 2020. 104142

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https:// doi. org/ 10. 5116/ ijme. 4dfb. 8dfd

Teitcher, J. E. F., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015).Detecting, preventing, and responding to “fraudsters” in internet research: Ethics and tradeoffs. Journal of Law, Medicine & Ethics, 43(1), 116–133. https://doi.org/10.1111/jlme.12200

The Cochrane Collaboration. (2011a). 9.5.2 identifying and measuring heterogeneity. In J. P. T. Higgins & S. Green (Eds.), Cochrane handbook for systematic reviews of interventions.

The Cochrane Collaboration. (2011b). Table7.7.a: Formulae for combining groups. In J. P. T. Higgins & S. Green (Eds.), Cochrane handbook for systematic reviews of interventions, Version 510.

Thompson, F. E., & Subar, A. F. (2017). Dietary assessment methodology. In AnnM. Coulston, Carol J. Boushey, Mario G. Ferruzzi, & Linda M. Delahanty (Eds.), Nutrition in the prevention and treatment of disease, 5–48). Academic Press.

Tinlin, L., Fini, N., Bernhardt, J., Lewis, L. K., Olds, T., & English, C. (2018). Best practice guidelines for the measurement of physical activity levels in stroke survivors: A secondary analysis of an observational study. International Journal of Rehabilitation Research, 41(1), 14–19. https://doi.org/10.1097/MRR.0000000000000253

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511819322 trials. BMJ, 366. https://doi.org/10.1136/bmj.l4898

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory Assessment. *Annual Review of Clinical Psychology*, *9*(1), 151–176. https://doi.org/10.1146/annurev-clinpsy-050212-185510

Turner-McGrievy, G. M., Yang, C.-H., Monroe, C., Pellegrini, C., & West, D. S. (2021). Is burden always bad? Emerging low-burden approaches to Mobile dietary self-monitoring and the role burden plays with engagement. *Journal of Technolog y in Behavioral Science*, *6*(3), 447–455. https:// doi. org/ 10. 1007/ s4134 7-021- 00203-9 University of Lübeck. (2025). ATI Scale. https:// ati-scale. org/ (accessed September 28, 2025).

U.S. Department of Agriculture, & U.S. Department of Health and Human Services. (2020). *Dietary Guidelines for Americans, 2020-2025. 9th Edition*. DietaryGuidelines.gov

Ullrich, A., Baumann, S., Voigt, L., John, U., & Ulbricht, S. (2021). Measurement reactivity of accelerometer-based sedentary behavior and physical activity in 2 assessment periods. Journal of Physical Activity and Health, 1(aop), 1–7. https://

Valenčič, E., Beckett, E., Collins, C. E., Koroušić Seljak, B., & Bucher, T. (2023). Snacktrack-an app-basedtool to assess the influence of digital and physical environments on snack choice. *Nutrients*, *15*(2), 349. https:// doi. org/ 10. 3390/ nu150 20349

Vanhelst, J., Béghin, L., Drumez, E., Coopman, S., & Gottrand, F. (2017). Awareness of wearing an accelerometer does not affect physical activity in youth. BMC Medical Research Methodology, 17(1), 1–6. https://doi.org/10.1186/s12874-017-0378-5

Vartanian, L. R., Herman, C. P., & Polivy, J. (2007). Consumption stereotypes and impression management: How you are what you eat. Appetite, 48(3), 265–277. https://doi.org/10.1016/j.appet.2006.10.008

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Walter, S. G., & Rack, O. (2009). Eine anwendungsbezogene Einführung in die Hierarchische Lineare Modellierung (HLM). In S. Albers, D. Klapper, U. Konradt, A. Walter, & J. Wolf (Eds.), *Methodik der empirischen Forschung* (pp. 381–396). Gabler Verlag. https://doi.org/10.1007/978-3-322-96406-9_25

Wansink, B., & Sobal, J. (2006). Mindless eating. *Environment and Behavior*, *39*(1), 106–123. https:// doi. org/ 10. 1177/ 00139 16506 295573

Webb, E., Campbell, D., Schwartz, R., & Sechrest, L. (2000). *Unobtrusive Measures, Revised Edition*. SAGE Publications, Inc. https://doi.org/10.4135/9781452243443

Webb, J., Baker, A., Palmer, T., Hall, A., Ahlquist, A., Darlow, J., Olaniyan, V., Horlock, R., & Stewart, D. (2022). The barriers and facilitators to physical activity in people with a musculoskeletal condition: A rapid review of reviews using the COM-B model to support intervention development. Public Health in Practice (Oxford, England), 3, 100250. https://doi.org/10.1016/j.puhip.2022.100250

Weinstein, N. D., & Sandman, P. M. (1992). A model of the precaution adoption process: Evidence from home radon testing. *Health Psychology*, *11*(3), 170–180. https://doi.org/10.1037/0278-6133.11.3.170

Weinstein, N. D., Rothman, A. J., & Sutton, S. R. (1998). Stage theories of health behavior: Conceptual and methodological issues. *Health Psychology*, *17*(3), 290–299. https://doi.org/10.1037/0278-6133.17.3.290

Wessel, D., Attig, C., & Franke, T. (2019). ATI-S – An ultra-short scale for assessing affinity for technology interaction in user studies. In F. Alt, A. Bulling, & T. Döring (Eds.), *Proceedings of mensch und computer 2019* (pp. 147–154). ACM. https:// doi.org/ 10. 1145/ 33407 64. 3340766

Whitton, C., Healy, J. D., Dhaliwal, S. S., Shoneye, C., Harray, A. J., Mullan, B. A., McVeigh, J. A., Boushey, C. J., & Kerr, D. A. (2023). Demographic and psychosocial correlates of measurement error and reactivity bias in a 4-d image-based mobile food record among adults with overweight and obesity. *British Journal of Nutrition*, *129*(4), 725–736. https://doi.org/10.1017/S0007114522001532

WHO. (2003). Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation. World Health Organization. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=284684

WHO. (2015). Guideline: Sugars intake for adults and children. Worl\d Health Organization. https://www.ncbi.nlm.nih.gov/books/NBK285537/

Wickström, G., & Bendix, T. (2000). The "Hawthorne effect"—What did the original Hawthorne studies actually show? *Scandinavian Journal of Work, Environment & Health*, *26*(4), 363–367.

Wilding, S., Conner, M., Sandberg, T., Prestwich, A., Lawton, R., Wood, C., Miles, E., Godin, G., & Sheeran, P. (2016). The question-behaviour effect: A theoretical and methodological review and meta-analysis. *European Review of Social Psychology*, *27*(1), 196–230. https://doi.org/10.1080/10463283.2016.1245940

Williams, S. L., & French, D. P. (2011). What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour—And are they the same? *Health Education Research*, *26*(2), 308–322. https://doi.org/10.1093/her/cyr005

Winkler, G., & Döring, A. (1995). Kurzmethoden zur Charakterisierung des Ernährungsmusters: Einsatz und Auswertung eines Food-Frequency-Fragebogens. Ernährungs-Umschau, 42(8), 289–291.

Winkler, G., & Döring, A. (1998). Validation of a short qualitative food frequency list used in several German large scale surveys. Zeitschrift Fur Ernahrungswissenschaft, 37(3), 234–241. https://doi.org/10.1007/PL00007377

Wood, C., Conner, M., Miles, E., Sandberg, T., Taylor, N., Godin, G., & Sheeran, P. (2016). The impact of asking intention or self-prediction questions on subsequent behavior: A meta-analysis. Personality and Social Psychology Review, 20(3), 245–268. https://doi.org/10.1177/1088868315592334

Wrzus, C., & Neubauer, A. B. (2022). Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538

Yang, C., Linas, B., Kirk, G., Bollinger, R., Chang, L., Chander, G., Siconolfi, D., Braxton, S., Rudolph, A., & Latkin, C. (2015). Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR mHealth and uHealth, 3(2), e67. https://doi.org/10.2196/mhealth.4344

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of Personality and Social Psychology, 9(2, Pt. 2), 1–27.https://doi.org/10.1037/h0025848

Zhu, X., & Haegele, J. A. (2019). Reactivity to accelerometer measurement of children with visual impairments and their family members. Adapted Physical Activity Quarterly, 36(4), 492–500. https://doi.org/10.1123/apaq.2019-0040

Zhu, X., Haegele, J., Wang, D., Zhang, L., & Wu, X. (2020). Reactivity to accelerometer measurement of youth with moderate and severe intellectual disabilities. Journal of Intellectual Disability Research, 64(9), 667–672. https://doi.org/10.1111/jir.12757

Ziesemer, K., König, L. M., Boushey, C. J., Villinger, K., Wahl, D. R., Butscher, S., Müller, J., Reiterer, H., Schupp, H. T., & Renner, B. (2020). Occurrence of and reasons for “missing events” in mobile dietary assessments: Results from three event-based ecological momentary assessment studies. JMIR mHealth and uHealth, 8(10), e15430. https://doi.org/10.2196/15430

Appendix

I. Paper 1

Appendix A – Search Strategies

EMBASE

1. (((physical activity or pedomet* or acceleromet* or smok* or "tobacco use" or alcohol* or drink* or diet* or food or snack* or eating behav* or dental or tooth* or teeth* or medication* or tablet* or sedentary behav*) and (measure* or assess*) and reactiv*) not (c-reactive or diethyl* or reactive oxygen)).ab. or (((physical activity or pedomet* or acceleromet* or smok* or "tobacco use" or alcohol* or drink* or diet* or food or snack* or eating behav* or dental or tooth* or teeth* or medication* or tablet* or sedentary behav*) and (measure* or assess*) and reactiv*) not (c-reactive or diethyl* or reactive oxygen)).ti.
2. limit 1 to yr="2008 - 2020"

Pubmed (incl. MEDLINE)

(physical activity[Title/Abstract] OR pedomet*[Title/Abstract] OR acceleromet*[Title/Abstract] OR smok*[Title/Abstract] OR Tobacco Use[Title/Abstract] OR alcohol*[Title/Abstract] OR drink*[Title/Abstract] OR diet*[Title/Abstract] OR food[Title/Abstract] OR snack*[Title/Abstract] OR eating behav*[Title/Abstract] OR dental[Title/Abstract] OR tooth*[Title/Abstract] OR teeth*[Title/Abstract] OR medication*[Title/Abstract] OR tablet*[Title/Abstract] OR sedentary behav*)[Title/Abstract] AND (measure*[Title/Abstract] OR assess*)[Title/Abstract] AND reactiv* NOT (c-reactive[Title/Abstract] OR diethyl*[Title/Abstract] OR reactive oxygen)[Title/Abstract]

PsycInfo

TI ((physical activity OR pedomet* OR acceleromet* OR smok* OR Tobacco Use OR alcohol* OR drink* OR diet* OR food OR snack* OR eating behav* OR dental OR tooth* OR teeth* OR medication* OR tablet* OR sedentary behav*) AND (measure* OR assess*) AND reactiv* NOT (c-reactive OR diethyl* OR reactive oxygen)) OR AB ((physical activity OR pedomet* OR acceleromet* OR smok* OR Tobacco Use OR alcohol* OR drink* OR diet* OR food OR snack* OR eating behav* OR dental OR tooth* OR teeth* OR medication* OR tablet* OR sedentary behav*) AND (measure* OR assess*) AND reactiv* NOT (c-reactive OR diethyl* OR reactive oxygen))

Web of Science Core Collection

TI=((physical activity OR pedomet* OR acceleromet* OR smok* OR Tobacco Use OR alcohol* OR drink* OR diet* OR food OR snack* OR eating behav* OR dental OR tooth* OR teeth* OR medication* OR tablet* OR sedentary behav*) AND (measure* OR assess*) AND reactiv* NOT (c-reactive OR diethyl* OR reactive oxygen)) OR AB=((physical activity OR pedomet* OR acceleromet* OR smok* OR Tobacco Use OR alcohol* OR drink* OR diet* OR food OR snack* OR eating behav* OR dental OR tooth* OR teeth* OR medication* OR tablet* OR sedentary behav*) AND (measure* OR assess*) AND reactiv* NOT (c-reactive OR diethyl* OR reactive oxygen))

Appendix B - List of data extracted

Study information

- First author
- Year of publication
- Journal name
- Geographical setting
- Number of studies
- Target behaviour(s) and description
- Target group(s)
- Study duration in days
- Study design (within or between subjects)
- Description of study design
- Description of assessment tool(s)

Information on participants

- Number of participants
- Specific characteristics of the sample

Information on results

- Description of condition and comparator (if available)
- Type of analysis conducted
- Moderators
- Control variables
- Type of effect size reported
- Reported effect size
- Analysis statistically significant?
- M, SD of condition and comparator (if available)
- Overall conclusion regarding measurement reactivity

Appendix C

Table 2. Effect sizes for the experimental studies. Sufficient detail to compute effect sizes could be obtained for 7 of the 10 experimental studies.

Study	Behaviour	Comparison			Cohen's d [a]
		Between or within participants	Condition in which less reactivity was expected	Condition in which more reactivity was expected	
Clemes et al. (2008) [b]	Physical activity: steps	within	covert	unsealed	0.84
Clemes and Parker (2009) [b]	Physical activity: steps	within	covert	sealed	0.18
Clemes and Parker (2009) [b]	Physical activity: steps	within	covert	unsealed	0.27
Clemes and Parker (2009) [b]	Physical activity: steps	within	covert	diary	0.65
Clemes and Parker (2009) [b]	Physical activity: steps	within	sealed	unsealed	0.16
Clemes and Parker (2009) [b]	Physical activity: steps	within	sealed	diary	0.41
Clemes and Parker (2009) [b]	Physical activity: steps	within	unsealed	diary	0.18

Study	Behaviour	Comparison			Cohen's d [a]
		Between or within participants	Condition in which less reactivity was expected	Condition in which more reactivity was expected	
Clemes and Deans (2012)	Physical activity: steps	within	covert	diary	N/A
Foley et al. (2011) [c]	Physical activity: activity counts	within	covert	unsealed	0.04
Foote et al. (2017)	Physical activity: activity counts	within	sealed	unsealed	N/A
Ho et al. (2013) [d]	Physical activity: activity counts	between	covert	unsealed	0.11
Ho et al. (2013) [d]	Physical activity: MVPA	between	covert	unsealed	0.28
McCarthy et al. (2015)	Smoking	between	low frequency	high frequency	N/A
Prewitt et al. (2013) [b]	Physical activity: steps	within	sealed	unsealed	0.06
Scott et al. (2014) [e]	Physical activity: steps	between	daily sealed	unsealed	-0.26
Scott et al. (2014) [e]	Physical activity: steps	between	weekly sealed	unsealed	0.42

Study	Behaviour	Comparison			Cohen's d [a]
		Between or within participants	Condition in which less reactivity was expected	Condition in which more reactivity was expected	
Vanhelst et al. (2017) [e]	Physical activity: activity counts	between	covert	sealed	0.36
Vanhelst et al. (2017) [e]	Physical activity: LPA	between	covert	sealed	0.38
Vanhelst et al. (2017) [e]	Physical activity: MPA	between	covert	sealed	0.30
Vanhelst et al. (2017) [e]	Physical activity: VPA	between	covert	sealed	0.20
Vanhelst et al. (2017) [e]	Physical activity: MVPA	between	covert	sealed	0.32
Vanhelst et al. (2017) [e]	Sedentary behaviour	between	covert	sealed	0.42

Note. [a] Positive values indicate an effect in the assumed direction. [b] Effect size calculated based on original dataset retrieved from the authors based on Borenstein et al. (2009). [c] Effect size provided by the authors on request. [d] Separate results for boys and girls were combined as recommended in https://handbook-5-1.cochrane.org/chapter_7/table_7_7_a_formulae_for_combining_groups.htm. [e] Cohen's d was calculated based on means and standard deviations provided in the publication based on Borenstein et al. (2009).

Appendix D

Figure 5. Funnel plot for all experimental studies investigating reactivity to measuring physical activity that could be included in the meta-analysis, adjusted using the trim and fill method.

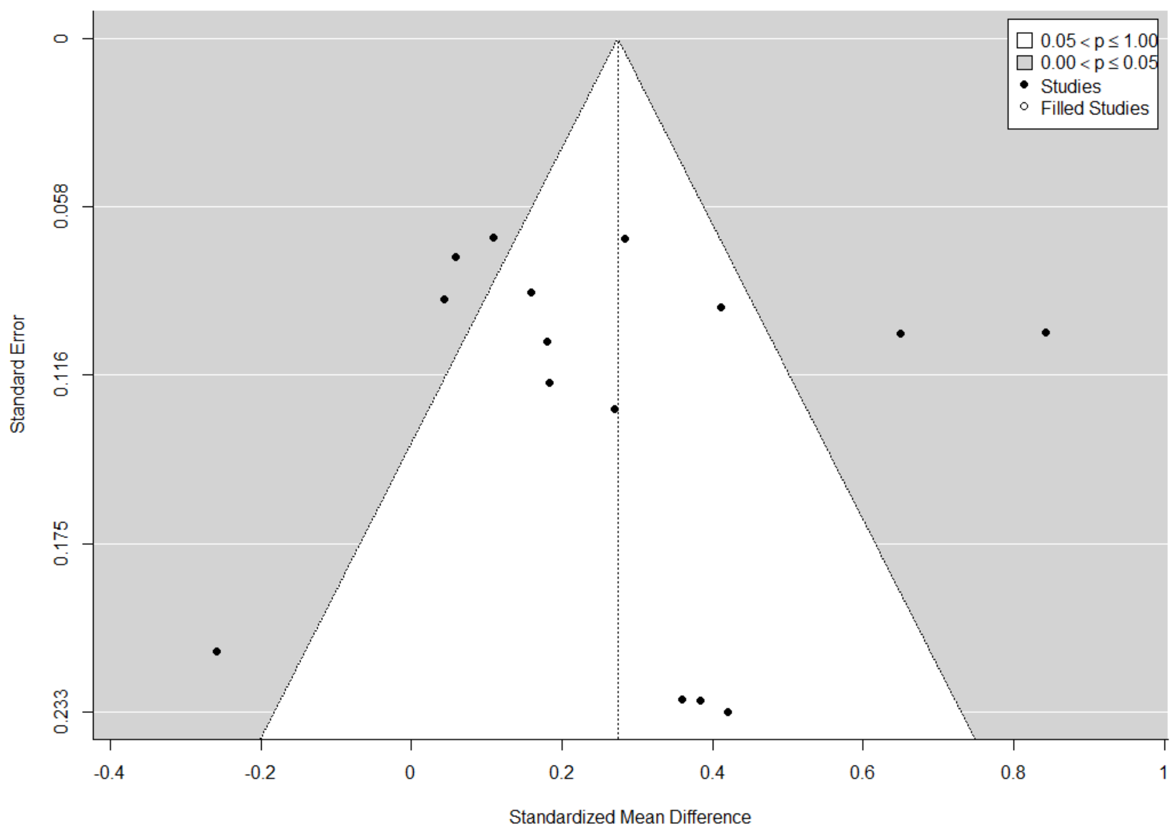


Figure 6. Funnel plot for all experimental studies investigating reactivity to measuring step counts, adjusted using the trim and fill method.

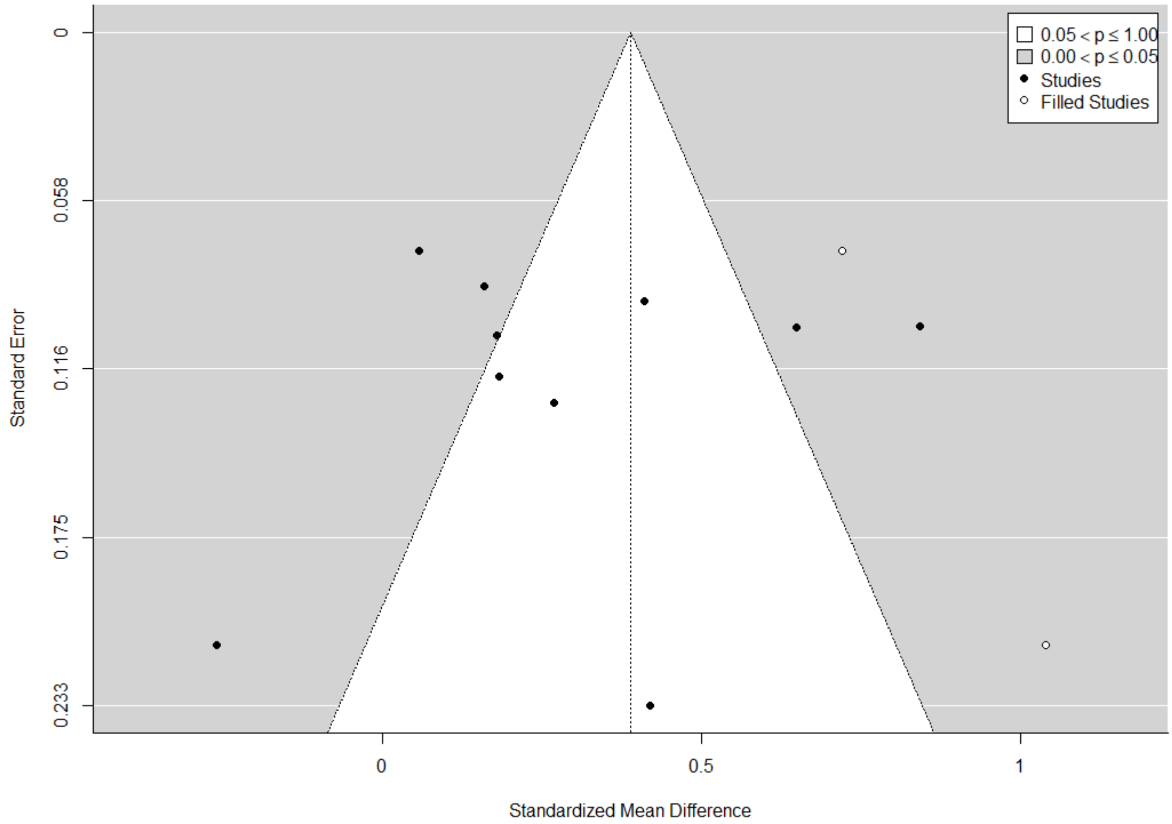


Table 3. Risk of bias assessment for randomised between-subject studies.

Study	Bias arising from the randomisation process	Bias due to deviations from the intended interventions (effect of assignment to intervention)	Bias due to missing outcome data	Bias in measurement of the outcome	Bias in selection of the reported result	Overall rating
Ho et al. (2013)	High	Some concerns	Low	Low	Some concerns	High
McCarthy et al. (2015)	High	Low	Low	Low	Some concerns	High
Scott et al. (2014)	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
Vanhelst et al. (2017)	Low	Some concerns	Low	Low	Some concerns	Some concerns

Table 4. Risk of bias assessment for studies using a within-subjects design.

	Appropriate cross-over design	Randomised treatment order	Carry-over effect	Unbiased data	Allocation concealment	Blinding	Incomplete outcome data	Selective outcome reporting	Other bias
Clemes et al. (2008)	Low	High	Unclear	Low	Low	Unclear	Unclear	Low	Unclear
Clemes and Parker (2009)	Low	High	Unclear	Low	Low	Unclear	Unclear	Low	Unclear
Clemes and Deans (2012)	Low	High	Unclear	Low	Low	Unclear	Low	Low	Unclear
Foley et al. (2011)	Low	Unclear	Unclear	Low	Unclear	Unclear	Low	Low	Unclear
Foote et al. (2017)	Low	High	Unclear	Low	Unclear	Unclear	Low	Low	Unclear
Prewitt et al. (2013)	Low	High	Unclear	Low	Unclear	Unclear	Low	Low	Unclear

Table 5. Risk of bias for observational studies.

Study	Were the criteria for inclusion in the sample clearly defined?	Were the study subjects and the setting described in detail?	Was the exposure measured in a valid and reliable way?	Were objective, standard criteria used for measurement of the condition?	Were confounding factors identified?	Were strategies to deal with confounding factors stated?	Were the outcomes measured in a valid and reliable way?	Was appropriate statistical analysis used?
Baumann et al. (2018)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cook et al. (2012)	Yes	No	Yes	Yes	Yes	Not clear	Not clear	Yes
Craig et al. (2010)	No	Yes	es	Yes	Yes	NA	Yes	Yes
Davis and Loprinzi (2016)	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Dössegger et al. (2014)	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Haegele et al. (2020)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hilgenkamp et al. (2012)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Labhart et al. (2020)	Yes	No	Not clear	No	No	Not clear	No	Not clear
Ling et al. (2011)	No	No	Yes	Yes	No	No	Yes	Yes
Ling and King (2015)	No	Yes	Yes	Yes	No	No	Yes	Yes
Klenk et al. (2019)	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Motl and Dlugonski (2011)	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Motl et al. (2012), Study 1	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Motl et al. (2012), Study 2	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Poulton et al. (2019)	Yes	Yes	Yes	No	Yes	Not clear	Yes	Yes
Sutton et al. (2014)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tinlin et al. (2018)	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Ullrich et al. (2021)	No	Yes	Yes	No	Yes	Yes	Yes	Yes

Yang et al. (2015)	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Zhu and Haegele (2019)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zhu et al. (2020)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

II. Paper 2

Appendix A – List of EMA Protocols with the respective mock-ups and descriptions

Table 2. Mock-up features description

EMA Protocol	Tracking Feature Combination	Nr. of screenshots	Nr. of words
1	Detailed description	3	91
2	Classification system	3	95
3	Photo recording	3	112
4	Classification system + Serving size	4	134
5	Food Database + Serving size	4	152
6	Photo recording + Detailed description	4	159
7	Food Database + Detailed description + Serving size	7	318
8	Photo recording +Food Database + Detailed description + Serving size	8	383

The mock-ups are listed based on their complexity indicated by the number of screenshots and words used to describe the reporting procedure.

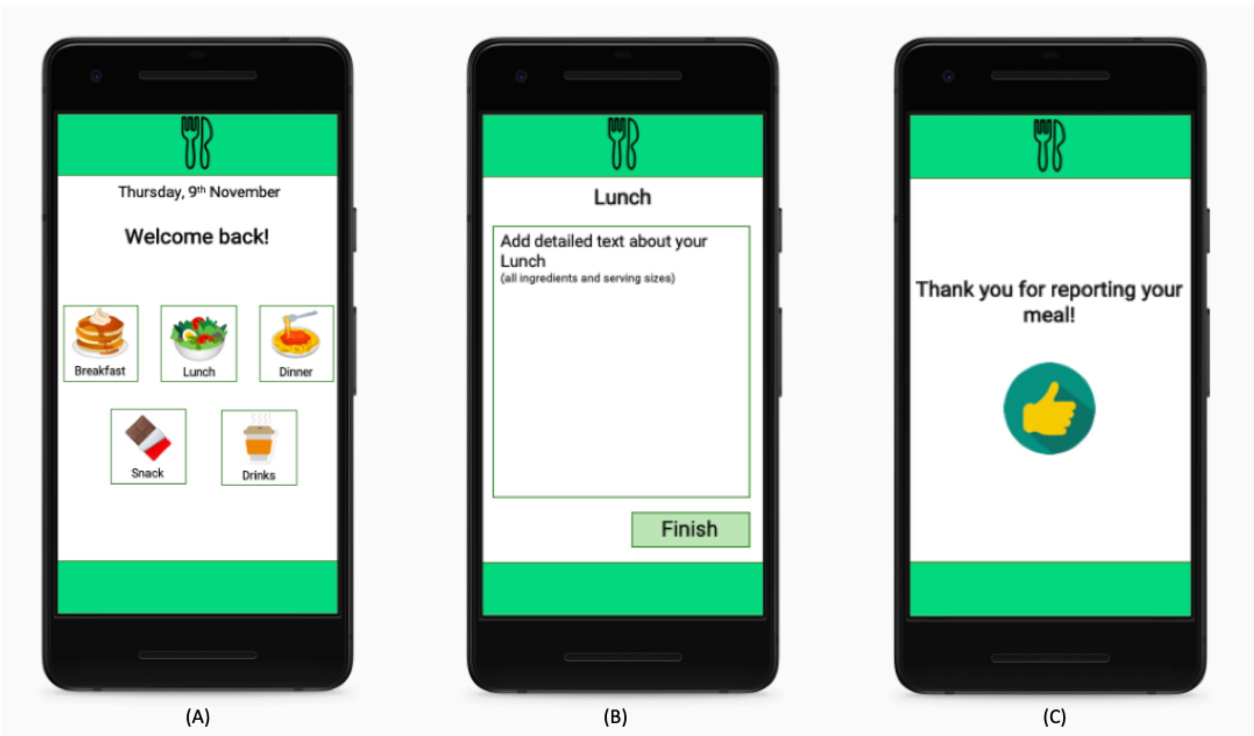
<u>Protocol 1: Detailed description (provide all main ingredients and serving sizes)</u>

Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

(A) Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by tapping one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) Next you can enter a brief description regarding your lunch (ingredients and serving size) in the text box. After adding all the components, you press "finish" to finish the recording.

(C) *You have reached the end of the recording processes. Thank you for reporting your meal!*



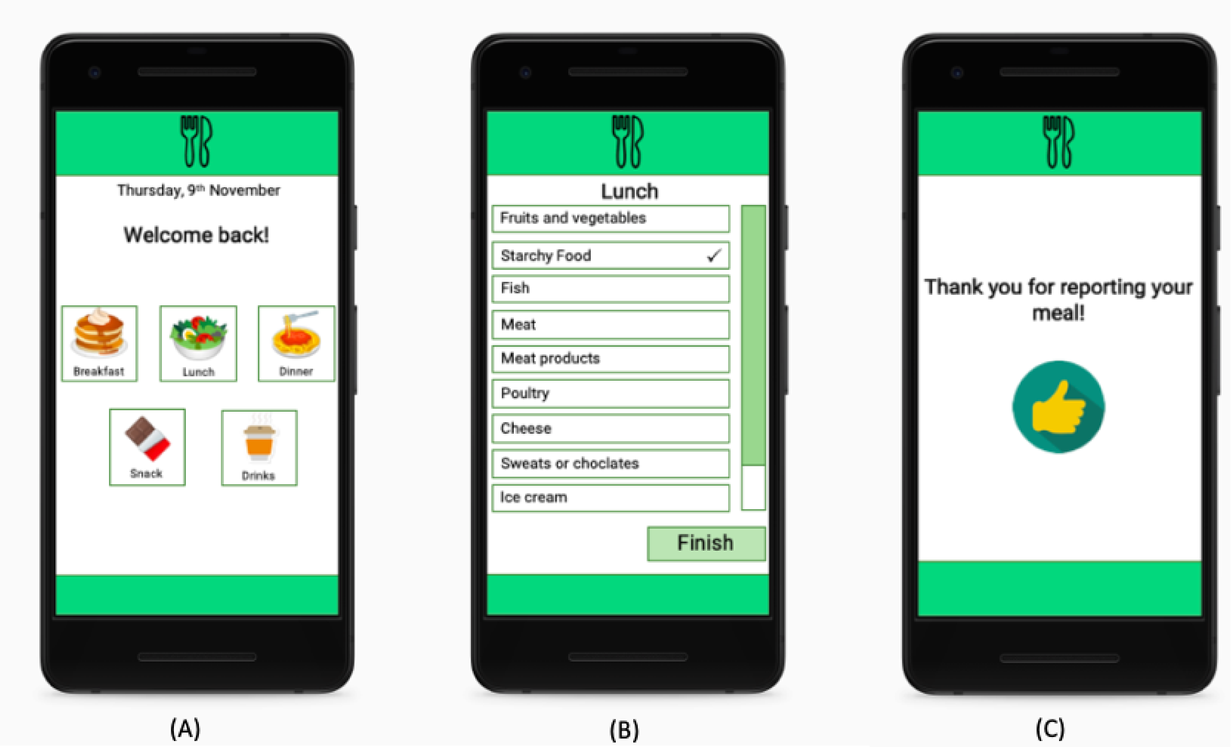
Protocol 2: Classification system

Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

(A) Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by pressing one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) Next, you see a list of 14 different classified food categories. By tapping the category, the OK (✓) button appears. After adding all categories, you tap on the "finish" button to complete the recording.

(C) *You have reached the end of the recording processes. Thank you for reporting your meal!*



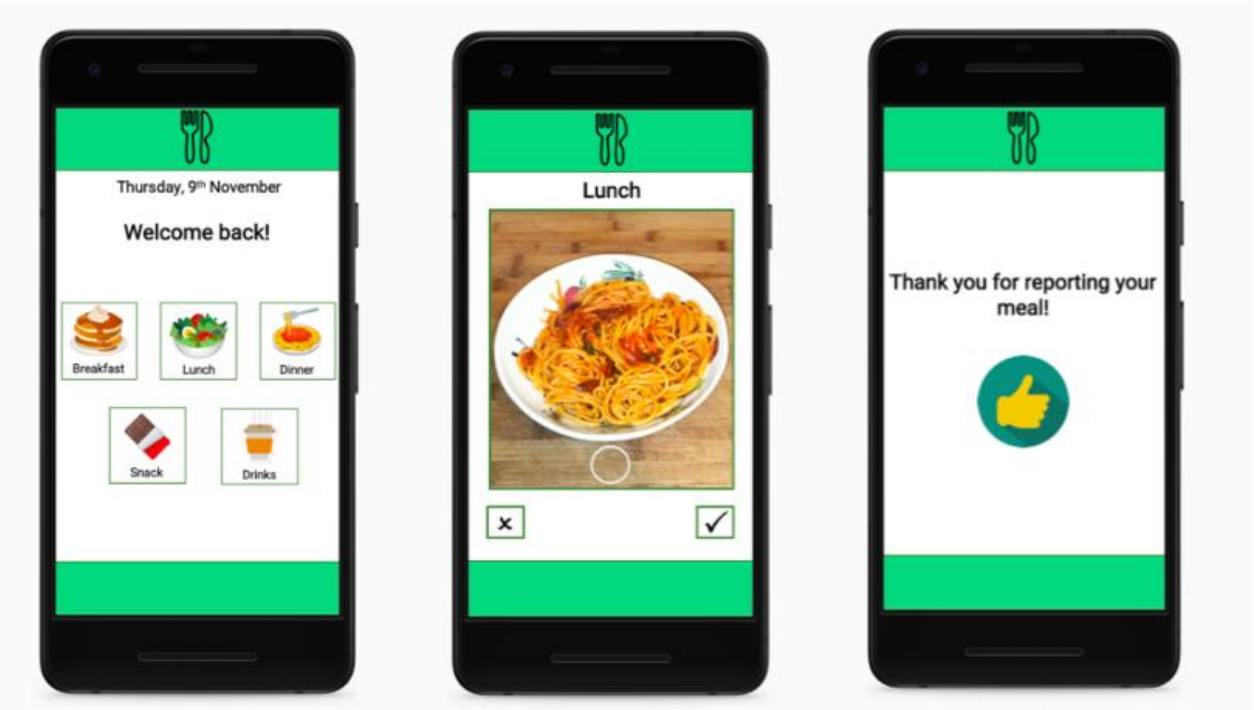
<u>Protocol 3: Photo-recording</u>

Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

(A) Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by pressing one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) Next, you take a photo of your food or drink. After taking the picture, you review it and accept it by tapping on the OK (✓) button to save the picture and to proceed. If you are not happy with the photo, you can discard it (×) and take another one.

(C) You have reached the end of the recording processes. Thank you for reporting your meal!



(A) (B) (C)

Protocol 4: Classification system + Serving size

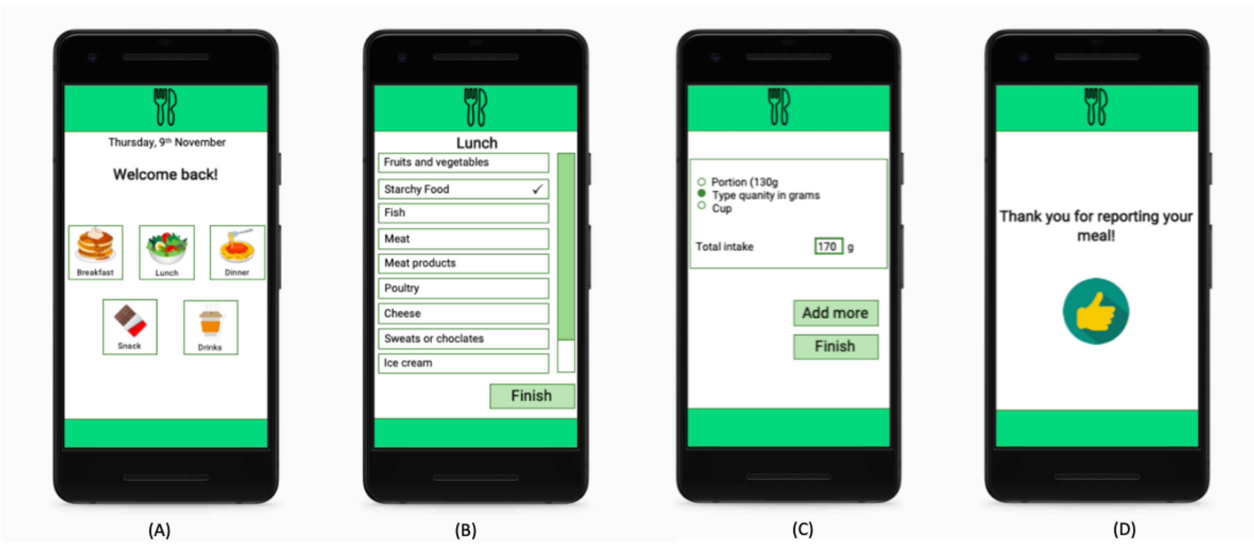
Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by pressing one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) Next, you see a list of 14 different classified food categories. After tapping one of these categories, you proceed automatically to add the serving size.

(C) You can select from three different serving size options: (a) portions; (b) type the quantity in grams; (c) cups. Next, you can tap the “add more” button to add more components until you have recorded all the ingredients of your meal or press “finish” to finish your recording.

(D) *You have reached the end of the recording processes. Thank you for reporting your meal!*



<u>Protocol 5: Food database + Serving size</u>

Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

(A) Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by pressing one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) You start to type in the search bar until the desired option appears. You select it by tapping on the respective line. You then tap on the "next" button and continue to the serving size of the food you just selected.

(C) Next you can select from three different serving size options: (a) portions; (b) type the quantity in grams; (c) cups. Next, you can tap the "add more" button to add more components until you have recorded all the ingredients of your meal or press "finish" to finish your recording.

(D) *You have reached the end of the recording processes. Thank you for reporting your meal!*

<u>Protocol 6: Photo recording + Detailed description</u>

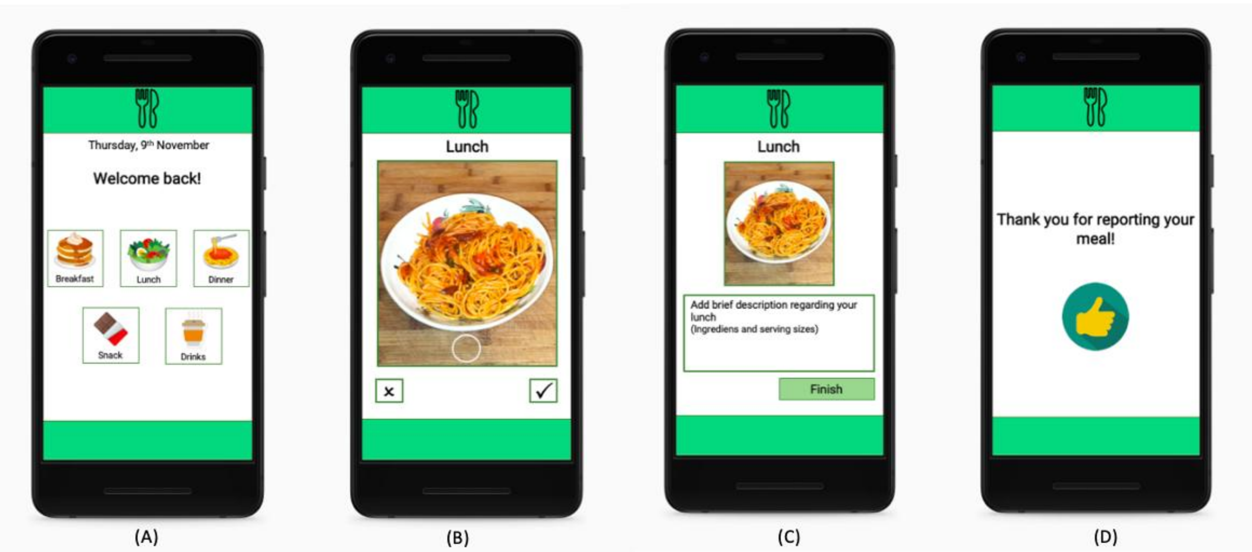
Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

(A) Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by tapping one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) Next, you take a photo of your food or drink. After taking the picture, you review it and accept it by tapping on the OK (✓) button to save the photo and to proceed. If you are not happy with the photo, you can discard it (✖) and take another one.

(C) Next, you see an overview of the food photo you just took and a Free text box where you can enter a brief description regarding your meal (ingredients and serving size). After adding all the components, you tap on the "finish" button to complete the recording.

(D) *You have reached the end of the recording processes. Thank you for reporting your meal!*



<u>Protocol 7: Food database + Detailed description + Serving size</u>

Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

(A) Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by tapping one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) Next, you add the components of the meal by tapping one of the two options *(1) search in database (2) add meal components as free text*

(C-1) If you tap on (1) you search for the foods in a food database. You start typing in the search bar until the correct option appears. You select it by tapping on the respective line. You then tap on the "next" button and continue to the serving size of the food you just selected.

(D) You can select from three different serving size options: (a) portions; (b) type the quantity in grams; (c) cups. Next, you tap the save button and you see an overview page (E) when you can repeat this process for every component of your meal.

(C-2) If you are unable to find a component of your meal, snack, or drink, in *search in database* you can tap "*add meal components as free text* "on the overview page (E) and enter the component in the text box. You then tap on the "next" button and continue to the serving size for the food you just selected (E).

(D) You can select from three different serving size options: (a) portions; (b) type the quantity in grams; (c) cups. Next, you tap the save button and you see an overview page (E) when you can repeat this process for every component of your meal.

Once you have recorded all components of your meal, you tap on the "*finish*" button in the overview page (E) to end the recording process.

(F) You have reached the end of the recording processes. Thank you for reporting your meal!

<u>Protocol 8: Photo recording + Food database + Detailed description + Serving size</u>

Imagine that you are using this app several times a day to record all your meals, snacks, and drinks.

A) Before you eat or drink, you open the app on your smartphone. First, you categorize the meal type by tapping one of five buttons: (1) breakfast, (2) lunch, (3) dinner, (4) snack, (5) drink. After tapping one of these categories, you proceed automatically.

(B) Next, you take a photo of your food or drink. After taking the photo, you review it and accept it by tapping on the OK (✓) button to save the picture and to proceed. If you are not happy with the photo, you can discard it (✖) and take another one.

(C) Next, you see an overview of the food photo you just took and add the components of the meal by tapping one of the two options *(1) search in database (2) add meal components as free text*

(D-1) If you tap on (1) you search for the foods in a food database. You start to type in the search bar until the correct option appears. You select it by tapping on the respective line. You then tap on the "next" button and continue to the serving size of the food you just selected.

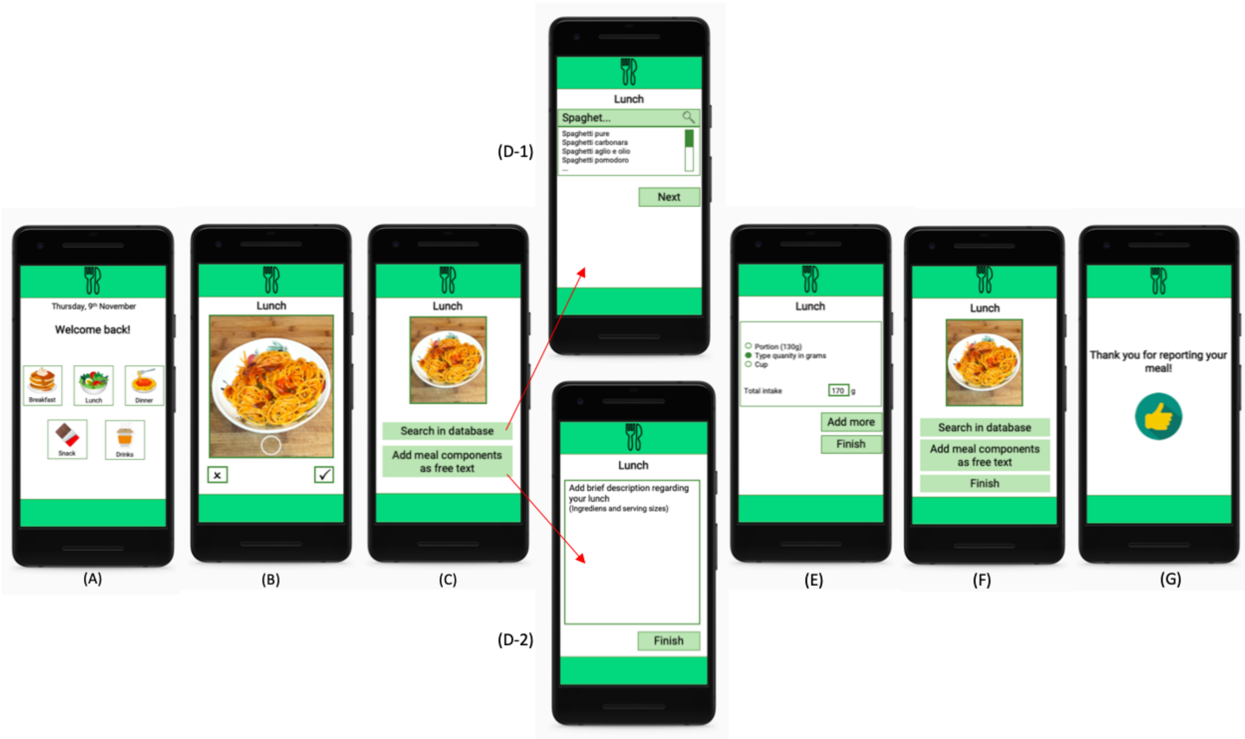
(E) You can select from three different serving size options: (a) portions; (b) type the quantity in grams; (c) cups. Next, you tap the save button and you see an overview page (F) when you can repeat this process for every component of your meal.

(D-2) If you are unable to find a component of your meal, snack, or drink, in *search in database* you can tap "*add meal components as free text* "on the overview page (F) and enter the component in the text box. You then tap on the "next" button and continue to the serving size for the food you just selected (E).

(E) You can select from three different serving size options: (a) portions; (b) type the quantity in grams; (c) cups. Next, you tap the save button and you see an overview page (F) when you can repeat this process for every component of your meal.

(F) Once you have recorded all components of your meal, you tap on the "*finish*" button in the overview page (F) to end the recording process.

(G) You have reached the end of the recording processes. Thank you for reporting your meal!


(D-1)
Lunch
Spaghet...
Spaghetti pure
Spaghetti carbonara
Spaghetti aglio e olio
Spaghetti pomodoro
Next
Thursday, 9th November
Welcome back!
Breakfast
Lunch
Dinner
Snack
Drinks
Lunch
Lunch
Search in database
Add meal components as free text
Lunch
Add brief description regarding your lunch
(Ingrediens and serving sizes)
Finish
Lunch
Portion (130g)
Type quanity in grams
Cup
Total intake
170 g
Add more
Finish
Lunch
Search in database
Add meal components as free text
Finish
Thank you for reporting your meal!
(A)
(B)
(C)
(D-2)
(E)
(F)
(G)

III. Paper 3

Supplementary Materials Ease of Change Study 1 + 2

List:

Table A.
Perceived ease of changing eating and physical activity behaviours. Means (M) and standard deviations (SD) for Studies 1 and 2

Aspects of Behavior	**Study 1** (*N*=435)		**Study 2** (*N*=496)	
	M	***SD***	**M**	***SD***
Eating Behavior:				
Eat no more than 300-600gr of meat/week	3.68	1.54	3.51	1.58
Eat healthier	3.48	1.10	4.12	1.36
Eat more snacks	3.46	1.23	3.90	1.45
Eat more overall	3.33	1.29	3.06	1.26
Eat 5 portions of fruit and vegetables	3.30	1.46	3.60	1.57
Eat additional meals	3.18	1.34	3.49	1.51
Eat larger portions	3.09	1.34	3.54	1.52
Eat fewer snacks	3.06	1.26	3.87	1.43
Eat no more than 50gr of added sugar/day	3.06	1.36	3.83	1.48
Eat less healthy	3.03	1.35	3.78	1.47
Eat smaller portions	2.84	1.21	4.15	1.29
Skip meals	2.77	1.42	3.77	1.55
Eat less overall	2.72	1.20	3.93	1.32
Physical Activity Behaviour:				
Spend less time standing	3.08	1.22	4.23	1.34
Be more physically active overall	3.00	1.19	3.79	1.54
Spend more time sitting	2.99	1.28	3.99	1.41

Engage in moderate/vigorous physical activity for 300min/week	2.97	1.35	3.07	1.68
Walk at least 10.000 steps/day	2.96	1.33	3.32	1.63
Spend more time standing	2.94	1.18	3.58	1.60
Be less physically active overall	2.79	1.38	3.72	1.51
Spend less time sitting	2.74	1.25	3.76	1.42

Table B.

Pairwise Comparison of Eating Behavior Aspects (Study 1)

Eating Behavior Aspects	2	3	4	5	6	7	8	9	10	11	12	13
1	<.001*	.030*	<.001*	<.001*	<.001*	.515	<.001*	<.001*	<.001*	<.001*	<.001*	<.001*
2	-	<.001*	<.001*	.055	<.001*	<.001*	.017*	.002*	.036*	.802	<.001*	.005*
3	-	-	.010*	<.001*	.042*	.356	<.001*	.003*	<.001*	<.001*	<.001*	.006*
4	-	-	-	<.001*	.370	<.001*	.167	.739	<.001*	.020*	<.001*	.773
5	-	-	-	-	<.001*	<.001*	<.001*	<.001*	.797	.011*	.017*	<.001*
6	-	-	-	-	-	.004*	.049*	.716	<.001*	.008*	<.001*	.716
7	-	-	-	-	-	-	<.001*	<.001*	<.001*	<.001*	<.001*	.001*
8	-	-	-	-	-	-	-	.188	<.001*	.195	<.001*	.220
9	-	-	-	-	-	-	-	-	<.001*	.004*	<.001*	.975
10	-	-	-	-	-	-	-	-	-	.076	.019*	<.001*
11	-	-	-	-	-	-	-	-	-	-	<.001*	.004*
12	-	-	-	--	-	-	-	-	-	-	-	<.001*

Note. **Eating related aspects**: 1= Eat less overall; 2= Eat more overall; 3= Eat smaller portions; 4= Eat larger portions; 5= Eat healthier; 6= Eat less healthy; 7= Skip meals; 8= Eat an additional meal; 9= Eat fewer snacks; 10= Eat more snacks; 11= Eat five portions of fruits and vegetables per day; 12= Eat no more than 300 to 600 grams of meat per week; 13= Eat no more than 50 grams of added sugar per day

*Correlation is significant at the 0.05 level (2-tailed)

Table C.
Pairwise Comparison of Physical Activity Aspects (Study 1)

Physical Activity Aspects	2	3	4	5	6	7	8
1	.048*	.471	.351	<.001*	.980	.582	.690
2	-	.096	<.001*	.599	.005*	.107	.096
3	-	-	.165	.001*	.529	.870	.758
4	-	-	-	<.001*	.180	.193	.257
5	-	-	-	-	.010*	.003*	.002*
6	-	-	-	-	-	.698	.793
7	-	-	-	-	-	-	.861

Note. **Physical Activity aspects**: 1= Be more physically active 2= Be less physically active; 3= Spend more time standing; 4= Spend less time standing; 5= Spend less time sitting; 6= Spend more time sitting; 7= Walk at least 10000 steps per day; 8= Engage in moderate to high intensity activities for at least 300 minutes per week
*Correlation is significant at the 0.05 level (2-tailed)

Correlations Tables Study 1

Study 1 (Social Comparison; Previous attempt to change eating behavior and PA)

Table D.

Pearson´s Correlation Social Comparison-Aspects of Eating Behavior (Study 1)

Aspects of Eating Behavior	**Social Comparison**			
	How healthy are their eating behaviors in comparison to others		How often they compare own eating behaviors to others	
	r	***p***	***r***	***p***
Eat Less Overall	.086	.075	-.019	.699
Eat more overall	.102	.034*	-.068	.157
Eat smaller portions	-.019	.698	.101	.036*
Eat larger portions	-.048	.318	.121	.012*
Eat healthier	.395	<.001*	.083	.085
Eat less healthy	-.330	.<.001*	.028	.560
Skip meals	-.034	.484	-.012	.804
Eat additional meals	-.037	.437	.090	.059
Eat fewer snacks	.203	<.001*	-.035	.472
Eat more snacks	-.085	.076	.074	.123
Eat 5 portions of fruit and vegetables	.339	<.001*	.012	.806
Eat no more than 300-600gr of meat/week	.141	.003*	.018	.709
Eat no more than 50gr of added sugar/day	.301	<.001*	.024	.611
	N=435		**N**=435	

Note. N= number of participants in the analysis
*Correlation is significant at the 0.05 level (2-tailed)

Table E.

Pearson´s Correlation Social Comparison-Aspects of Physical Activity (Study 1)

Aspects of Physical Activity	**Social Comparison**			
	How PA they think they are in comparison to others		How often they compare own PA levels to others	
	r	***p***	***r***	***p***
Be more physically active	.066	.172	.339	<.001*
Be less physically active	-.039	.413	-.374	<.001*
Spend more time standing	.093	.053	.143	.003*
Spend less time standing	.002	.973	-.144	.003*
Spend less time sitting	.055	.255	.226	<.001*
Spend more time sitting	-.022	.641	-.256	<.001*
Walk at least 10000 steps per day	.093	.053	.343	<.001*
Engage in moderate to high intensity activities for at least 300 minutes per week	.126	.008*	.598	<.001*
	N=435		**N**=435	

Note. N= number of participants in the analysis
*Correlation is significant at the 0.05 level (2-tailed)

Table F.

Pearson´s Correlation Previous attempts to change behavior-Aspects of Eating Behavior (Study 1)

Aspects of Eating Behavior	**Previous Attempts to change Eating Behavior**							
	Improve Eating Behavior		Start a formal healthy eating program		Start healthy eating program (< 3 days)		Start healthy eating program (≥ 3 days)	
	r	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Eat Less Overall	-.053	.283	.070	.155	.055	.268	.025	.608
Eat more overall	.154	.002*	.014	.781	.044	.375	.026	.602
Eat smaller portions	-.072	.147	.054	.275	-.058	.238	-.053	.288
Eat larger portions	.114	.021*	-.038	.437	.014	.776	.011	.831
Eat healthier	-.034	.487	.058	.233	.041	.412	.024	.634
Eat less healthy	.131	.008*	-.041	.408	.020	.690	.038	.442
Skip meals	.009	.864	.078	.113	.006	.907	.089	.071
Eat additional meals	.105	.034*	.109	.025*	-.002	.968	.065	.188
Eat fewer snacks	-.040	.425	.059	.229	-.092	.061	-.019	.700
Eat more snacks	.063	.203	.089	.070	.019	.698	-.022	.652
Eat 5 portions of fruit and vegetables	-.086	.084	.003	.959	.047	.338	-.021	.674
Eat no more than 300-600gr of meat/week	-.043	.385	-.092	.059	-.085	.084	-.009	.853
Eat no more than 50gr of added sugar/day	-.042	.396	-.083	.090	.014	.771	.014	.781
	N=435		**N**=435		**N**=435		**N**=435	

Note. N= number of participants in the analysis

*Correlation is significant at the 0.05 level (2-tailed)

Table G.

Pearson´s Correlation Previous attempts to change behavior-Aspects of Physical Activity (Study 1)

Aspects of Physical Activity	**Previous Attempts to change Physical Activity Levels**							
	Improve PA Level		Start a formal PA program		Start PA program (< 3 days)		Start PA program (≥ 3 days)	
	r	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Be more physically active	.059	.243	.066	.183	-.059	.235	.019	.703
Be less physically active	.106	.036*	-.039	.430	-.059	.230	-.015	.762
Spend more time standing	-.005	.925	.037	.457	.079	.111	.024	.627
Spend less time standing	-.057	.245	-.044	.369	.005	.915	.025	.617
Spend less time sitting	.067	.177	.043	.382	-.010	.841	-.042	.398

Spend more time sitting	-.027	.591	-.006	.907	.078	.113	.084	.091
Walk at least 10000 steps per day	-.077	.120	.097	.049*	.007	.891	.005	.915
Engage in moderate to high intensity activities for at least 300 minutes per week	.073	.139	.071	.151	.057	.253	.050	.316
	N=435		**N**=435		**N**=435		**N**=435	

Note. PA = Physical Activity, N= number of participants in the analysis
*Correlation is significant at the 0.05 level (2-tailed)

Study 2:

Table H.
Pairwise Comparison of Eating Behavior Aspects (Study 2)

Eating Behavior Aspects	2	3	4	5	6	7	8	9	10	11	12	13
1	.101	<.001*	<.001*	.004*	.084	.025*	<.001*	.315	.711	<.001*	<.001*	.162
2	-	<.001*	<.001*	<.001*	.979	.916	<.001*	.328	.078	.040*	.007*	.589
3	-	-	<.001*	.587	<.001*	<.001*	<.001*	<.001*	.006*	<.001*	<.001*	<.001*
4	-	-	-	<.001*	.002*	.024*	.425	<.001*	<.001*	.527	.763	.003*
5	-	-	-	-	<.001*	<.001*	<.001*	<.001*	.018*	<.001*	<.001*	<.001*
6	-	-	-	-	-	.885	<.001*	.363	.094	.071	.009*	.633
7	-	-	-	-	-	-	.003*	.235	.160	.079	.006*	.510
8	-	-	-	-	-	-	-	<.001*	<.001*	.230	.811	<.001*
9	-	-	-	-	-	-	-	-	.758	.001*	<.001*	.603
10	-	-	-	-	-	-	-	-	-	.001*	<.001*	.451
11	-	-	-	-	-	-	-	-	-	-	.307	.004*
12	-	-	-	--	-	-	-	-	-	-	-	<.001*

Note. **Eating related aspects**: 1= Eat less overall; 2= Eat more overall; 3= Eat smaller portions; 4= Eat larger portions; 5= Eat healthier; 6= Eat less healthy; 7= Skip meals; 8= Eat an additional meal; 9= Eat fewer snacks; 10= Eat more snacks; 11= Eat five portions of fruits and vegetables per day; 12= Eat no more than 300 to 600 grams of meat per week; 13= Eat no more than 50 grams of added sugar per day
*Correlation is significant at the 0.05 level (2-tailed)

Table I.
Pairwise Comparison of Physical Activity Aspects (Study 2)

Physical Activity Aspects	2	3	4	5	6	7	8
1	.522	.057	<.001*	.567	.005*	<.001*	<.001*
2	-	.052	<.001*	.741	.005*	<.001*	<.001*
3	-	-	<.001*	.005*	<.001*	<.001*	<.001*
4	-	-	-	<.001*	<.001*	<.001*	<.001*
5	-	-	-	-	.014*	<.001*	<.001*
6	-	-	-	-	-	<.001*	<.001*
7	-	-	-	-	-	-	<.001*

Note. **Physical Activity aspects**: 1= Be more physically active 2= Be less physically active; 3= Spend more time standing; 4= Spend less time standing; 5= Spend less time sitting; 6= Spend more time sitting; 7= Walk at least 10000 steps per day; 8= Engage in moderate to high intensity activities for at least 300 minutes per week
*Correlation is significant at the 0.05 level (2-tailed)

Correlations Tables Study 2

Study 2 (Social Comparison; Previous attempt to change eating behavior and PA)

Table J.

Pearson´s Correlation Social Comparison-Aspects of Physical Activity (Study 2)

Aspects of Physical Activity	**Social Comparison**			
	How PA they think they are in comparison to others		How often they compare own PA levels to others	
	r	***p***	***r***	***p***
Be more physically active	.54	<.001*	.04	.389
Be less physically active	-.28	<.001*	.01	.826
Spend more time standing	.45	<.001*	.06	.211
Spend less time standing	-.17	<.001*	.02	.603
Spend less time sitting	.39	<.001*	.01	.783
Spend more time sitting	-.18	<.001*	.05	.221
Walk at least 10000 steps per day	.55	<.001*	.07	.098
Engage in moderate to high intensity activities for at least 300 minutes per week	.59	<.001*	.12	.007*
	N=507		**N**=507	

Note. N= number of participants in the analysis
*Correlation is significant at the 0.05 level (2-tailed)

Table K.

Pearson´s Correlation Social Comparison-Aspects of Eating Behavior (Study 2)

Aspects of Eating Behavior	**Social Comparison**			
	How healthy are their eating behaviors in comparison to others		How often they compare own eating behaviors to others	
	r	***p***	***r***	***p***
Eat Less Overall	.28	<.001*	-.06	.145
Eat more overall	-.02	.682	.08	.065
Eat smaller portions	.25	<.001*	.002	.968
Eat larger portions	-.07	.103	.08	.081
Eat healthier	.48	<.001*	.04	.343
Eat less healthy	-.26	.<.001*	-.01	.841
Skip meals	-.03	.488	-.05	.276
Eat additional meals	-.06	.203	.08	.088
Eat fewer snacks	.28	<.001*	.05	.911
Eat more snacks	-.08	.057	.02	.700
Eat 5 portions of fruit and vegetables	.29	<.001*	.06	.193
Eat no more than 300-600gr of meat/week	.20	<.001*	.06	.182
Eat no more than 50gr of added sugar/day	.27	<.001*	.01	.758
		N=510		**N**=510

Note. N= number of participants in the analysis
*Correlation is significant at the 0.05 level (2-tailed)

Table L.

Pearson´s Correlation Previous attempts to change behavior-Aspects of Eating Behavior (Study 2)

Aspects of Eating Behavior	Previous Attempts to change Eating Behavior					
	Start a formal healthy eating program		Start healthy eating program (< 3 days)		Start healthy eating program (≥ 3 days)	
	r	***p***	***r***	***p***	***r***	***p***
Eat Less Overall	.12	.007*	.08	.083	.09	.039*
Eat more overall	.03	.938	-.03	.459	-05	.215
Eat smaller portions	.08	.068	.06	.141	.06	.145
Eat larger portions	.02	.596	-.02	.740	-.04	.427
Eat healthier	.13	.003*	.11	.010*	.10	.031*
Eat less healthy	.02	.720	-.05	.257	-.04	.393
Skip meals	.04	.371	.02	.721	.01	.748
Eat additional meals	.01	.928	-.05	.281	-.07	.100
Eat fewer snacks	.07	.113	.05	.276	.05	.279
Eat more snacks	-.10	.022*	-.07	.109	-.09	.050
Eat 5 portions of fruit and vegetables	.03	.499	.03	.522	.01	.838
Eat no more than 300-600gr of meat/week	.04	.402	.01	.787	.02	.711
Eat no more than 50gr of added sugar/day	.10	.030*	.09	.051	.11	.015*
	N=507		**N**=507		**N**=507	

Note. N= number of participants in the analysis

*Correlation is significant at the 0.05 level (2-tailed)

Table M.

Pearson´s Correlation Previous attempts to change behavior-Aspects of Physical Activity (Study 2)

Aspects of Physical Activity	Previous Attempts to change Physical Activity Levels							
	Improve PA Level		Start a formal PA program		Start PA program (< 3 days)		Start PA program (≥ 3 days)	
	r	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Be more physically active	-.06	.186	.05	.280	-.02	.724	.01	.889
Be less physically active	-.05	.305	-.06	.172	-.10	.024*	-.06	.194
Spend more time standing	-.05	.238	.04	.356	-02	.629	-.04	.391
Spend less time standing	.03	.476	.01	.901	-.06	.182	-.03	.473
Spend less time sitting	-.02	.702	.03	.492	.01	.821	.01	.878
Spend more time sitting	-.03	.469	.03	.487	-.06	.188	-.02	.689

Walk at least 10000 steps per day	-.02	.710	.05	.261	.08	.058	.03	.480
Engage in moderate to high intensity activities for at least 300 minutes per week	-.01	.921	.09	.051	.10	.020*	-.01	.970
	N=507		**N**=507		**N**=507		**N**=507	

Note. N= number of participants in the analysis
*Correlation is significant at the 0.05 level (2-tailed)

IV. Paper 4

Supplementary Material - Table of Contents

Supplementary Tables

Supplementary analyses

Exploratory research question

We exploratively compare the perceived complexity of the EMA study app, active information processing, and perceived changes in eating-related cognitions, intentions, and behaviors due to the tracking process between this sample and the sub-sample that received a comparable app mockup (using photo-based and detailed description as food intake tracking features) in Allmeta et al. (2026) to provide an indicator for whether the results of this mockup – based study mirror ratings obtained in a real-life context.

Measures

Ease of Change. The perceived ease of change of 13 aspects of eating behaviour was assessed (see Table 1 for the full list of items) (Allmeta, Arigo, et al., 2026). These included both aspects of behaviour that are usually considered "desired changes" i.e., *eat healthier, eat fewer snacks*, as well as aspects that are usually considered "undesired changes" i.e., *eat less healthy, skip meals*. Additionally, 3 out of the 13 aspects of eating behaviour ("*eating 5 portions of fruits and vegetables per day"* (World Health Organization, 2003) *"eating no more than 300 to 600 grams of meat per week"* (Clinton et al., 2020)*, "eating no more than 50 grams of added sugar per day"* (World Health Organization, 2015), were based on international recommendations and guidelines. The terms added sugar was further defined in a footnote[5]; see the materials in OSF.

Participants were asked to rate the 13 aspects on a six-point scale from (1) *would be very difficult for me* to (6) *would be very easy for me*. Means and standard deviations are

[5] Added sugar refers to any sugar that is added to food by the manufacturer, cook or consumer, as well as sugars that are naturally contained in honey, syrups and fruit juices.

reported in Table 1. When rating, they were explicitly asked to compare each behaviour listed to what they were currently doing.

Table 1
Perceived ease of changing of eating behavior aspects. Means (M) and Standard deviations (SD)

Aspects of eating behavior	**Sample (*N* = 80)**	
	M	***SD***
Skip meals	2.69	1.44
Eat less overall	2.86	0.87
Eat smaller portions	3.20	1.06
Eat fewer snacks	3.21	1.30
Eat no more than 50gr of added sugar/day	3.34	1.52
Eat 5 portions of fruit and vegetables	3.40	1.42
Eat larger portions	3.56	1.05
Eat additional meals	3.64	1.21
Eat healthier	3.74	1.04
Eat less healthy	3.76	1.39
Eat more overall	3.85	1.03
Eat more snacks	4.11	1.32
Eat no more than 300-600gr of meat/week	4.84	1.65

Perceived Complexity. Participants were asked to rate their experience in using the app to record their food intake for 8 consecutive days with six items on a 5-point Likert scale (1-*strongly disagree* to 5-*strongly agree*). For instance, they rated whether recording meals and snacks with the study app was detailed, took a long time or required a lot of input (Allmeta, Sutton, et al., 2026b)

Eating related cognition, intention, and behaviour. Eating-related cognitions were assessed with 6 items. For example, participants indicated whether using the study app made them think about the healthiness of the meal or become aware of how much they ate.

Intentions to change behaviour were assessed with 5 items. For instance, participants rated whether using the study app made them want to eat less or want to eat fewer snacks. Anticipated changes in eating behaviour were assessed through 5 items. Amongst others, participants rated whether using the study app would make them eat less or eat healthier. The same 5-point Likert scale was used in these three measures (1-*strongly disagree* to 5- *strongly agree*).

The scales perceived complexity, active information processing, eating-related cognition, intention and behaviour were used to assess the same constructs in a similar context in a previous study (Allmeta, Sutton, et al., 2026). The reliability of the listed scales was calculated as internal consistency (Cronbach´s alpha). Items reflecting the complexity of the study app (Cronbach´s α = .39) did not show a good internal consistency, while items reflecting eating related cognitions (Cronbach´s α = .81), intentions (Cronbach´s α = .81), and behaviour (Cronbach´s α = .71) showed good internal consistencies with values above .7 (Tavakol & Dennick, 2011).

Statistical Analysis

Means and standard deviations are reported in Table 2 for (1) perceived EMA study app complexity, (2) active information processing, (3) perceived changes in eating related cognition, intention and behaviour for both the sample of the current study and the subsample from Allmeta, Sutton, et al., (2026) that received the description of a highly similar app. The ratings were compared using independent sample *t*-tests for each (sub-)scale.

Table 2
Exploratory RQ. Mean, SD and Test Statistic

Aspects	Sample (N = 80)		Sub-Sample (N = 42)		Test Statistic			
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*Cohen´s d*
Active Information Processing	2.80	0.64	3.39	0.81	-4.43	120	<.001*	-.84

Perceived EMA App Complexity	2.61	0.66	2.56	0.74	.31	106	.756	.06
Eating-related Cognition	3.78	0.67	3.80	0.79	-.09	120	.929	-.02
Eating-related Intention	2.47	0.81	3.24	0.82	-4.98	120	<.001*	-.95
Eating-related Behavior	2.27	0.67	3.05	0.78	-5.80	120	<.001*	-1.11

Results and discussion

Independent samples *t*-tests were conducted to compare the mean scores of perceived EMA app complexity, active information processing, eating related cognitions, intentions and behaviors between this studies sample and the sub-sample of a previous study (see Allmeta et al., 2026 for context). The results indicated that there is a statistically significant difference between the sample and sub-sample regarding active information processing, eating related intention and behavior. In all three significant cases, the mockup-based sub-sample reported higher mean scores than the real-life sample (see Table 4) with effect sizes ranging from small to large (Cohen´s *d* = .02 - 1.11). There is no significant difference regarding the perceived complexity of the EMA study app and eating related cognitions.

The interpretation of these results suggests a hypothetical bias, where participants in the mockup group may have overestimated their likely engagement due to the lack of real-world barriers (Ajzen et al., 2024). The large effect sizes for intention and behavior suggests an intention-behavior gap, a phenomenon where high motivation in controlled or simulated settings fails to translate into action during real-life implementation (Sheeran & Webb, 2016). While participants can accurately anticipate the complexity of a tool from a mockup, they appear to struggle with forecasting the cognitive effort (active information processing) required for daily EMA compliance. This suggests that mockup-based pilot studies may provide overly optimistic projections of intervention efficacy – but potentially also for perceived burden as indicated by the divergence in experienced vs anticipated AIP – as real-world ecological burdens may impact the behavioral intentions formed in a hypothetical scenario.

References

Ajzen, I., Brown, T. C., & Carvajal, F. (2004). Explaining the Discrepancy between Intentions and Actions: The Case of Hypothetical Bias in Contingent Valuation. *Personality and Social Psychology Bulletin*, *30*(9), 1108–1121. https://doi.org/10.1177/0146167204264079

Allmeta, A., Arigo, D., & König, L. M. (2026). Easy or difficult? Investigating perceived ease of changing eating and physical activity behaviors. *Applied Psychology: Health and Well-Being*, *18*(1), e70124. https://doi.org/https://doi.org/10.1111/aphw.70124

Allmeta, A., Sutton, S., & König, L. M. (2026). The same, only different: Smartphone-based dietary Ecological Momentary Assessment tools vary in complexity, usability and active information processing. *British Journal of Health Psychology*, *31*(1), e70057. https://doi.org/10.1111/bjhp.70057

WHO. (2003). *Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation*. World Health Organization. **https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=284684**

WHO. (2015). *Guideline: Sugars intake for adults and children*. Worl\d Health Organization. **https://www.ncbi.nlm.nih.gov/books/NBK285537/**

Sheeran, P., & Webb, T. L. (2016). The Intention–Behavior gap. *Social and Personality Psychology Compass*, *10*(9), 503–518. https://doi.org/10.1111/spc3.12265

Supplementary tables

Table 3
Perceived ease of changing of eating behavior aspects. Means (M) and Standard deviations (SD)

Aspects of eating behavior	**Sample (*N* = 80)**	
	M	***SD***
Skip meals	2.69	1.44
Eat less overall	2.86	0.87
Eat smaller portions	3.20	1.06
Eat fewer snacks	3.21	1.30
Eat no more than 50gr of added sugar/day	3.34	1.52
Eat 5 portions of fruit and vegetables	3.40	1.42
Eat larger portions	3.56	1.05
Eat additional meals	3.64	1.21
Eat healthier	3.74	1.04
Eat less healthy	3.76	1.39
Eat more overall	3.85	1.03
Eat more snacks	4.11	1.32
Eat no more than 300-600gr of meat/week	4.84	1.65

Table 4
Average Consumption of the outcome variables per day. Mean and SD

Outcome / **Day**	Food Intake (g)		Fruits & Vegetables (g)		Snacks (g)		Snacks (count)		Skipped meals (count)	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
1	829.47	408.77	296.78	227.38	118.43	145.84	1.35	1.17	0.27	0.45
2	745.36	418.31	254.97	215.43	108.72	117.71	1.4	1.24	0.23	0.45
3	741.00	354.25	255.04	193.11	109.2	149.21	1.21	1.15	0.21	0.54
4	569.59	372.81	209.56	190.84	120.57	137.07	1.38	1.4	0.26	0.49
5	664.43	386.46	261.29	192.9	106.44	132.67	1.29	1.21	0.32	0.52
6	620.13	354.99	212.19	186.8	97.09	128.41	0.99	1.07	0.2	0.46
7	673.94	346.73	244.47	178.13	81.89	125.9	0.98	1.02	0.27	0.61
8	644.03	379.07	219.13	205.1	94.44	155.29	1.29	1.38	0.25	0.51

Table 5.
Random Intercept Model (fixed effects) – RQ1 Model A (time as continues variable)

RQ	1a - Number of snack/day				1b - Amount of snack/day					1c - Amount of fruit & vegetable / day				
	b	SE	z	*p*	b	SE	t	df	*p*	b	SE	t	df	*p*
Intercept	0.17	0.09	1.90	.058	119.30	11.64	10.25	234.49	< .001*	276.59	17.69	15.64	185.78	< .001*
Day (0-7)	-0.03	0.02	-2.21	.027*	-4.16	2.04	-2.04	570.76	.042*	-8.12	2.79	-2.91	570.57	.004*
Weekend	-0.01	0.08	-0.03	.973	-0.26	10.69	-0.02	571.14	.980	-15.84	14.63	-1.08	571.06	.280

RQ	2 – Number of skipped meals/ day				2 – Amount of food consumed/day				
	b	SE	z	*p*	b	SE	t	df	*p*
Intercept	-1.85	0.20	-9.2	< .001*	781.72	36.65	21.33	87.77	< .001*
Day (0-7)	0.01	0.04	0.09	.927	-23.47	5.76	-4.07	80.96	< .001*
Weekend	0.28	0.17	1.68	.092	-55.36	27.19	-2.04	532.47	0.012

Table 6.
Random Intercept Model – RQ1 Model B (time as categorical variable)

RQ	1a - Number of snack/day				1b - Amount of snack/day					1c - Amount of fruit & vegetable / day				
	b	SE	z	*p*	b	SE	t	df	*p*	b	SE	t	df	*p*
Intercept	0.10	0.12	0.80	.421	95.11	15.22	6.25	467.16	< .001*	219.53	22.13	9.92	367.70	< .001*
Day 1	0.06	0.14	0.45	.655	23.76	18.74	1.27	564.52	.205	77.85	25.49	3.05	564.39	.002*
Day 2	0.11	0.14	0.78	.446	15.05	18.80	0.80	564.52	.424	37.40	25.56	1.46	564.36	.144
Day 3	-0.04	0.14	-0.29	.773	15.97	18.85	0.85	564.53	.397	38.08	25.64	1.49	564.40	.138
Day 4	0.10	0.15	0.71	.480	30.54	19.89	1.54	564.67	.125	-3.02	27.05	-0.11	564.49	.911
Day 5	0.06	0.15	0.38	.702	18.07	20.80	0.87	564.78	.386	50.98	28.29	1.80	564.57	.072
Day 6	-0.24	0.15	-1.63	.103	4.41	18.96	0.23	564.55	.816	-4.02	25.78	-0.16	564.41	.876
Day 7	-0.27	0.15	-1.81	.069	-12.89	18.80	-0.69	564.15	.493	24.88	25.56	0.97	564.12	.339

Weekend	-0.06	0.09	-0.66	.509	-9.05	13.00	-0.70	565.95	.487	-12.39	17.69	-0.70	565.38	.484

Table 7.
Random Intercept Model (fixed effects) – Cross-level Interactions RQ 3, 4, 5. Time as continues variable x moderators

Outcome / Moderator	Snacks (count/day)				Snacks (g/day)					Fruits and Vegetables (g/day)				
	b	SE	z	*p*	b	SE	t	df	*p*	b	SE	t	df	*p*
RQ3														
Intercept	0.16	0.09	1.70	.089	118.697	11.85	10.02	225.51	< .001*	277.10	17.85	15,52	180.43	< .001*
Time (day 0-7)	-0.03	0.02	-2.11	.035	-4.05	2.07	-1.96	557.00	.051	-8.06	2.82	-2.86	557.00	.004*
AIP (gmc=0)	0.24	0.15	1.66	.097	30.13	18.13	1.66	202.51	.098	11.51	27.49	0.42	164.34	.676
Weekend	0.01	0.08	0.09	.929	0.12	10.86	0.01	558.23	.991	-15.75	14.82	-1.06	557.88	.288
Time * Active Information Processing	-0.03	0.03	-1.15	.252	-3.27	3.27	-1.01	557.00	.317	9.08	4.46	2.04	556.99	.042*
RQ4														
Intercept	0.18	0.09	1.90	.057	118.88	11.78	10.09	228.25	< .001*	279.65	17.84	15.68	181.12	< .001*
Time (day 0-7)	-0.03	0.02	-2.14	.032*	-4.02	2.06	-1.95	562.76	.052	-8.54	2.81	-3.04	562.57	.003*
Intention to change eating behavior (gmc=0)	0.07	0.10	0.71	.477	5.63	13.26	0.42	204.94	.672	-0.35	20.19	-0.02	165.04	.986
Weekend	0.01	0.08	0.08	.937	0.98	10.81	0.09	563.43	.927	-16.84	14.75	-1.14	563.03	.254
Time * Intention	-0.01	0.02	-0.26	.797	0.85	2.39	0.35	562.16	.723	4.45	3.26	1.37	562.12	.173
RQ5														
Intercept	0.23	0.11	2.05	.041*	124.04	14.12	8.78	223.54	< .001*	276.54	21.51	12.86	178.27	< .001*
Time (day 0-7)	-0.06	0.02	-2.87	.004*	-6.38	2.49	-2.56	570.05	.011*	-6.53	3.42	-1.91	569.79	.056

Previous Experience (0,1)	-0.15	0.19	-0.81	.418	-14.08	24.04	-0.59	206.66	.559	-0.08	36.79	-0.01	166.64	.998
Weekend	-0.01	0.08	-0.07	.943	-0.56	10.68	-0.05	570.44	.958	-15.63	14.64	-1.07	570.04	.286
Time * Experience	0.06	0.03	1.86	.063	6.68	4.33	1.54	569.44	.123	-4.79	5.93	-0.81	569.34	.420