MASTERARBEIT | MASTER'S THESIS

Titel | Title

Measurement Reactivity in Experienced versus Novice Physical Activity Trackers

verfasst von | submitted by

Johanna Fleck BSc BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Univ.-Prof. Dr. Laura Maria König B.Sc. M.Sc.

Abstract

Measurement reactivity, or changes in behaviour, cognition, and emotion solely due to measurement, may occur in the assessment of physical activity and bias findings. The present study aimed to investigate two potential mechanisms of measurement reactivity, namely familiarity with recording physical activity and the salience of researcher observation. Recruitment targeted people with and without prior experience with personal fitness apps and trackers; and participants were assigned one of two salience conditions, which more or less emphasised researcher observation. Thirty-one participants (aged $M = 22.87$ years, $SD = 7.62$; 74.19% female, 25.81% male) wore an accelerometer (movisens Move 4) for 14 days recording daily steps taken. Multilevel models showed no significant interactions between study day and experience ($b = -60.87$, $SE = 86.05$, $df = 402.02$, $t = -0.71$, $p = .48$), nor between study day and salience ($b = 117.28$, $SE = 86.03$, $df = 402.03$, $t = 1.36$, $p = .174$), nor a significant three-way interaction ($b = 266.63$, $SE = 147.74$, $df = 222.99$, $t = 1.81$, $p = .073$). However, since no overall measurement reactivity in terms of a decline in steps over the study period could be attested ($b = -25.95$, $SE = 75.74$, $df = 427.83$, $t = -0.34$, $p = .732$ in the last model including the three-way interaction), the ability to draw comprehensive conclusions about its potential mechanisms based on this study is limited. Further research successfully inducing measurement reactivity is required to clarify its mechanisms.

Zusammenfassung

Messreaktivität oder die Veränderung von Verhalten, Kognitionen und Emotionen allein durch die Messung kann bei der Erhebung von körperlicher Aktivität auftreten und Ergebnisse verzerren. Ziel der vorliegenden Arbeit war es, zwei potenzielle Mechanismen von Messreaktivität zu untersuchen, nämlich die Vertrautheit mit dem Aufzeichnen von körperlicher Aktivität und die Salienz der Beobachtung durch Forschende. Es wurden gezielt Personen mit und ohne Vorerfahrung mit eigenen Fitness-Apps und -Trackern rekrutiert, und Versuchspersonen wurden einer von zwei Salienzbedingungen zugeordnet, in der die Beobachtung durch das Forschungsteam mehr oder weniger betont wurde. Einunddreißig Versuchspersonen (Alter M = 22.87 Jahre, SD = 7.62; 74.19% weiblich, 25.81% männlich) trugen einen Bewegungssensor (movisens Move 4) zur Aufzeichnung der täglich zurückgelegten Schritte für 14 Tage. In den berechneten Multilevelmodellen zeigten sich weder signifikante Interaktionen zwischen dem Studientag und der Vorerfahrung (b = -60.87, SE = 86.05, df = 402.02, t = -0.71, p = .48) noch dem Studientag und der Salienzbedingung (b = 117.28, SE = 86.03, df = 402.03, t = 1.36, p = .174) und auch keine dreifache Interaktion (b = 266.63, SE = 147.74, df = 222.99, t = 1.81, p = .073). Da allerdings keine generelle Messreaktivität im Sinne einer signifikanten Abnahme der Schrittzahlen über den Studienzeitraum festgestellt wurde (b = -25.95, SE = 75.74, df = 427.83, t = -0.34, p = .732 im letzten Modell mit dreifacher Interaktion), ist die Aussagekraft dieser Studie über deren potenzielle Mechanismen stark beschränkt. Es bedarf weiterer Forschung, die erfolgreich Messreaktivität induziert, um deren Mechanismen zu klären.

Table of Contents

Theoretical Background

Empirical psychological research generally employs participants, about whose behaviour, cognition, and emotion conclusions are intended to be drawn based on observation. Often, it is implicitly assumed that the mere act of participating in an empirical study does not in itself impact the participants (French et al., 2021; French & Sutton, 2010; McCambridge, Kypri, & Elbourne, 2014). However, such a belief is recently being interrogated more thoroughly in discussion surrounding *research participation effects*, which refer to a range of potential ways in which participating in research may affect the participants (McCambridge, Kypri, & Elbourne, 2014; McCambridge, Witton, & Elbourne, 2014). Previously, this issue had been touched upon in investigations falling under the umbrella of the *Hawthorne effect*, which has lately been criticised for being too ill-defined a concept to have much utility (Chiesa & Hobbs, 2008; McCambridge, Witton, & Elbourne, 2014). Empirical evidence exists that the awareness of being studied, for example by being made to answer questions on one's behaviour or being directly observed, may indeed affect a participant's behaviour (McCambridge, Witton, & Elbourne, 2014). However, there is difficulty in affirming the overarching conditions under which such an effect may or may not arise nor its potential size because of the large heterogeneity of studies included and the plethora of potential factors involved in research participation effects (McCambridge, Witton, & Elbourne, 2014). Considering this difficulty in drawing conclusions about a phenomenon so broad, the present study is dedicated to investigating one specific aspect of research participation effects, namely those associated with the act of measurement.

The following section will thus introduce the central phenomenon this study intends to investigate, namely *measurement reactivity*, defining the term, and outlining why and when it can be problematic. Afterwards, this thesis will turn to the second key concept, *physical activity*, addressing different ways of measuring it and their implications. Then, these two concepts will be brought together in a review of the current state of research on measurement reactivity in physical activity measured with objective motion sensors, which will clarify the research gap the present study attempted to fill.

Measurement Reactivity

When the act of measurement causes changes in the people being assessed, this may be termed measurement reactivity (French & Sutton, 2010). More specifically, the attempt to measure participants' behaviour, thoughts and emotions as part of a research study may affect the very constructs under investigation. Without measurement, such changes would not occur

(French et al., 2021). Generally, it has not yet been well established which areas of psychological research might be affected by measurement reactivity effects, with two exceptions: relatively convincing evidence exists that measurement reactivity may appear when posing participants questions on their behaviour, also known as the *question-behaviour effect*, and when measuring physical activity with objective wearable monitors (French et al., 2021). In the context of health behaviours, in which most measurement reactivity research to date has been conducted, there appears a tentative pattern for the direction of behaviour change associated with measurement reactivity to be related to social norms, i.e. behaviour commonly deemed socially desirable or healthy increases while socially undesirable or unhealthy behaviour decreases (French et al., 2021). It has been proposed (French et al., 2021; French & Sutton, 2010) that measurement procedures may bring about change in participants' behaviour because they inadvertently tap into *behaviour change techniques*, which describe those elements of behaviour change interventions responsible for effecting change (Michie et al., 2015), an example of which would be self-monitoring.

Although unintended changes in research participants due to measurement are generally undesirable, the existence of measurement reactivity per se does not necessarily lead to notable bias in study results (French et al., 2021; French & Sutton, 2010). Depending on the exact context, circumstances and strength of the effect, however, bias due to measurement reactivity may be a marked issue for psychological research. Three scenarios in which measurement reactivity effects may be particularly problematic for experimental studies were first identified by French and Sutton (2010). First, measurement reactivity and an experimental intervention may be confounded, for example if similar mechanisms underlie both effects. This could be the case if an intervention aiming to increase physical activity by fostering self-monitoring measured this behaviour via pedometers, and these pedometers by themselves led to heightened self-monitoring. Second, there may be an interaction between the effects of the measurement and the intervention. If the measurement itself already increased a behaviour like physical activity, which presumably cannot be raised endlessly, then any intervention would not be able to produce any further effect. Third, difficulty may arise in the measurement of a given attribute affected by measurement reactivity. An overall elevation of mean scores due to measurement reactivity could make ceiling effects more likely, potentially leading to a reduction in power to detect differences, depending on the statistical methods in question. Adding to the work done by French and Sutton (2010), a further three scenarios in which measurement reactivity is particularly likely to cause bias were identified by French et al. (2021) more recently. In particular, differential measurement between groups in a trial, such that some groups may be measured more often, or with different measures, than others, could lead to systematic between-group differences induced by the measurement protocol. Furthermore, considering that the measurement may be burdensome for participants, differential measurement between groups may also lead to differential attrition, especially if an intervention is already taxing, thus leading to a much heavier load for the intervention group. Finally, the measurement procedures could affect interactions with participants, for instance, if one group had more frequent contact with health care professionals than another group and thus received superior care. This outline of different scenarios particularly prone to measurement reactivity is intended to illustrate the varied intricacies of when studies indeed may or may not allow the intended interpretation of results.

Despite the fact that first recommendations on dealing with measurement reactivity in trials have been produced, these have been based mainly on indirect evidence and there remain many aspects of this phenomenon which are currently not well understood (French et al., 2021). As such, the relevance of more research into the scope of measurement reactivity and its various potential mechanisms is clear. One reason that measurement reactivity is a particularly pressing concern in psychological research is that small effects should be accepted as the norm in this area (Götz et al., 2022). Thus, if effect sizes are generally not especially pronounced, then also small measurement reactivity effects could lead to meaningful bias. Furthermore, whether or not measurement reactivity impacts results in such a way that they are not valid and accurate is an especially important consideration when it comes to factors related to human health, which is a concern particularly pertinent to the field of health psychology. In this context, physical activity is a crucial health behaviour, with well-documented associations to both physical health (Ekelund et al., 2019; Jakicic et al., 2019; Paluch et al., 2022; Reiner et al., 2013; World Health Organization [WHO], 2020) and mental health (Bendau et al., 2024; Firth et al., 2020; WHO, 2020). Physical inactivity is widespread (Strain et al., 2020, 2024; WHO, 2018) and leads to considerable health burden (Ammar et al., 2023) and economic burden (Ding et al., 2016) worldwide. It is thus especially important that physical activity can be measured precisely so that interventions and recommendations concerning this health behaviour can be as accurate as possible. Consequently, this study is dedicated to the examination of measurement reactivity concerning physical activity.

Physical Activity

In the following section, the focus of this thesis will shift towards physical activity, the second key concept in the present study. A brief consideration of what physical activity entails and how it can be measured is aimed to enable an informed understanding of the methods utilised and the particular scope of this research project.

Definition and Key Aspects of Physical Activity

According to Strath et al. (2013), most commonly used definitions of physical activity can be traced back to Caspersen et al. (1985), who proposed that the term may subsume “any bodily movement produced by skeletal muscles that results in energy expenditure” (p. 126). The World Health Organisation (WHO), for example, also utilises an equivalent description of the term (WHO, 2018, 2020). Physical activity is a complex behaviour which may be further specified and subdivided according to a range of aspects. One such differentiation is based on whether physical activity is structured or incidental, with the former referring to an intentional activity carried out with the aim of improving physical fitness, also known as *exercise*, and the latter describing any other less deliberate kind of physical activity (Caspersen et al., 1985; Strath et al., 2013). Moreover, physical activity may be described according to four different dimensions (Ainsworth et al., 2015; Strath et al., 2013). First, frequency refers to how often physical activity is performed, for example in times per day or per week. Second, the duration of physical activity denotes the length of a given bout of physical activity within a certain time frame, for example in minutes or hours per day or week. Third, the intensity of physical activity is related to the amount of energy expended through bodily movements. This encompasses both the relative intensity of the activity, which may vary based on a person’s cardiorespiratory fitness, and the absolute intensity, defined by the kind of physical activity performed and commonly expressed in units of metabolic equivalents (METs), where 1 MET describes the energy demand of sitting quietly (Strath et al., 2013). Ordinarily, physical activity intensity is separated into light, moderate, and vigorous physical activity, including sedentary behaviour on the lower end of the activity spectrum. Sedentary behaviour may be classified as any activity up to a maximum of 1.5 METs, light physical activity as ranging from 1.6 to 2.9 METs, moderate physical activity as 3 to 5.9 METs, and vigorous physical activity as 6 METs or more (Strath et al., 2013). The fourth dimension refers to the mode or type of physical activity, for instance walking or cycling (Ainsworth et al., 2015; Strath et al., 2013). Lastly, different physical activity domains may be distinguished, which typically include occupational, household or domestic, transportation and leisure time physical activity (Ainsworth et al., 2015; Strath et al., 2013). Considering these various aspects involved in physical activity reveals how complex a behaviour it is. As a result, physical activity can be challenging to assess (Skender et al., 2016).

Physical Activity Measurement

Subjective Versus Objective Measures of Physical Activity. There is a large number of ways in which physical activity can be measured. These may be broadly organised into two groups: subjective and objective measures of physical activity (Ainsworth et al., 2015; Strath et al., 2013). Examples for subjective measures are physical activity questionnaires, which generally ask participants to retrospectively estimate the frequency, duration, and intensity of physical activity within a given time frame, for instance the past week or month, and physical activity logs and diaries, in which participants continuously record all physical activity they perform (Ainsworth et al., 2015; Strath et al., 2013). Wearable motion sensors like pedometers and accelerometers, physiological measures like heart-rate monitors (Ainsworth et al., 2015; Strath et al., 2013) as well as measures of energy expenditure based on indirect calorimetry or doubly-labelled water, which rely on indicators of human metabolic processes, in addition to direct observation may all be subsumed under the umbrella of objective physical activity measures (Strath et al., 2013). Combining multiple objective measures for an even more detailed picture of physical activity is also possible (Ainsworth et al., 2015; Strath et al., 2013).

Subjective measures have historically been the chief way of assessing physical activity, and much of our current understanding of the relationship between physical activity and health is based on studies employing this type of physical activity assessment (Ainsworth et al., 2015). Completing questionnaires especially carries a minimal burden for the respondent, comes with a high versatility in use, is cost-effective, and generally accepted as a method by both researchers and participants (Ainsworth et al., 2015). However, due to methodological limitations, subjective measures of physical activity are increasingly superseded by objective measures (Ainsworth et al., 2015). This development may be illustrated by the fact that from 2006 to 2016, the percentage of lifestyle physical activity interventions employing objective measures has risen from 4.4% to 70.6% (Silfee et al., 2018). Especially retrospective questionnaires may be prone to recall biases or social desirability biases (Burchartz et al., 2020). Concerning psychometric properties, a study of 64 adults and children found that accelerometer-determined physical activity was more reliable, valid, and stable over two separate measurement weeks in comparison to a physical activity questionnaire or diary, though quality criteria were generally higher for moderate compared to vigorous physical activity and some differences between the age groups existed (Fiedler et al., 2021). Furthermore, a systematic umbrella review summarising 63 review articles on the methodological effectiveness of a range of physical activity measures in adults also indicated that objective monitors may exhibit higher reliability and validity than subjective measures, leading the authors to advocate for the use of objective physical activity measures, while acknowledging that differences in study design, data processing and statistical analysis between studies complicate comparability between specific measurement types and tools (Dowd et al., 2018).

Interestingly, comparisons of subjective and objective measures of physical activity have suggested that there is limited agreement in physical activity estimates based on these two methods, with error that does not follow clearly predictable patterns (Ainsworth et al., 2015; Skender et al., 2016). This finding may be related to the strengths and potential blind spots of each measurement technique, with questionnaires being able to assess physical activity more easily over long time periods and various types of physical activity, while objective methods may not be equally sensitive to some types of physical activity, like stationary activity for example, but can record activity continuously and are not affected by biases associated with self-report (Skender et al., 2016). Still, subjective measures should not be considered obsolete since they offer the unique opportunities in studying specific aspects of physical activity like types, domains and the social, psychological and environmental context of physical activity (Sattler et al., 2021). The most comprehensive assessments of physical activity may be achieved by a combination of different measurement methods (Burchartz et al., 2020; Chen & Bassett, 2005; Pedišić & Bauman, 2015; Reichert et al., 2020; Sattler et al., 2021; Skender et al., 2016). Deliberations on which tool to use should always be guided by considerations of the advantages and disadvantages relevant to a given research context (Burchartz et al., 2020; Sattler et al., 2021).

The preceding general overview of physical activity measurement methods is intended to provide the necessary broader context for the following section, in which the technology of and key methodological considerations on objective physical activity monitors will be expounded on. After all, this study aims to explore measurement reactivity effects in this precise context.

Objective Physical Activity Monitoring. Two of the most frequently used physical activity monitors are pedometers and accelerometers (Ainsworth et al., 2015; Strath et al., 2013). Pedometers were designed mainly to measure ambulatory activity based on the number of steps walked (Ainsworth et al., 2015; Strath et al., 2013). Modern pedometers typically employ microelectromechanical systems and algorithm-based processing to identify a step (Ainsworth et al., 2015). Accelerometers determine bodily accelerations, measured in gravitational units g or m/s^2, in one to three planes (Ainsworth et al., 2015; Chen & Bassett, 2005; Strath et al., 2013). They also use microelectromechanical systems (Ainsworth et al., 2015). Apart from the number of steps taken, accelerometers may also identify a range of physical activity indicators, for instance, units of energy expenditure like METs or physical activity intensity categories (Burchartz et al., 2020; Strath et al., 2013). Although it has been previously suggested that Japanese pedometers in particular are the most accurate instruments for the determination of steps in an analysis of the large-scale 2005-2006 National Health and Nutrition Examination Survey (NHANES), where the accelerometer-derived steps were deemed unrealistically high (Tudor-Locke et al., 2009), a more recent umbrella review indicated high levels of criterion validity for both activity monitors, with accelerometers even providing more accurate estimates of steps than pedometers (Dowd et al., 2018).

The basic technology of accelerometer-based physical activity measurement is described by Chen and Bassett (2005). During acceleration, a seismic mass inside the accelerometer will deform or compress a piezoelectric element, causing a build-up of electrical charge, which then can be transformed into an electrical output signal proportional to the amount of acceleration. This basic setup, qualified for detecting the acceleration in one specific plane, is duplicated orthogonally for accelerometers capable of detecting acceleration along multiple axes. Due to the physical properties of the piezoelectric element, reliable identification works best for dynamic events, as the electrical charge will be depleted over time if only static acceleration is present, a process also termed leakage. The processing of the raw electrical output signal involves several different steps, finally leading to accelerometer counts as the primary physical activity outcome (Chen & Bassett, 2005), which is frequently further transformed into a number of different physical activity indicators, typically by the application of algorithms and intensity thresholds (Burchartz et al., 2020; Strath et al., 2013). For each point in the data treatment process, there is a host of alternative choices a researcher could reasonably make, with important implications for the interpretability of any given study and the comparability between different studies (Ainsworth et al., 2015; Burchartz et al., 2020; Chen & Bassett, 2005; Dowd et al., 2018; Pedišić & Bauman, 2015; Strath et al., 2013).

In addition to the data processing decisions made, there are further aspects of the design of accelerometer-based studies that can impact the physical activity estimates yielded. The exact model of the monitor in use may be one such aspect, both in terms of different accelerometer brands and different generations of one model (Ainsworth et al., 2015; Burchartz et al., 2020; Pedišić & Bauman, 2015; Strath et al., 2013). Moreover, the use of uniaxial versus three-axial vector-magnitude data, even stemming from the same device, may also impact physical activity estimates (Keadle et al., 2014). Furthermore, physical activity estimates may vary depending on the body part the monitor is placed at like the hip, wrist, ankle, or trunk (Ainsworth et al., 2015; Mora-Gonzalez et al., 2022). The exact type of physical activity examined also matters, with the most accurate assessment attained for ambulatory activities (Ainsworth et al., 2015; Pedišić & Bauman, 2015), but reduced accuracy for sedentary and light-intensity physical activity (Ainsworth et al., 2015; Mora-Gonzalez et al., 2022) as well as the higher end of the intensity spectrum (Chen & Bassett, 2005). Activities which accelerometers have an especially limited ability to accurately capture include cycling, weight lifting (Ainsworth et al., 2015; Chen & Bassett, 2005; Matthews, 2005; Pedišić & Bauman, 2015; Silfee et al., 2018; Skender et al., 2016), household chores (Ainsworth et al., 2015; Matthews, 2005) and swimming, since many motion sensors are not waterproof (Chen & Bassett, 2005; Pedišić & Bauman, 2015; Silfee et al., 2018; Skender et al., 2016). The context of physical activity is also relevant here, with higher accuracy achieved in laboratory conditions compared to free-living activity (Ainsworth et al., 2015; Burchartz et al., 2020), which is an important consideration if validation studies for a specific device are mainly laboratory-based (Burchartz et al., 2020).

An additional factor with the potential to influence physical activity estimates is accelerometer wear time (Pedišić & Bauman, 2015), with two relevant facets. First, it may make a difference what kind of wear time protocol is implemented within a study, i.e. whether the accelerometer should be worn by participants during all hours of the day or only during waking hours (Burchartz et al., 2020). A 24-hour protocol has been previously recommended (Burchartz et al., 2020; Migueles et al., 2017), but should be weighed against ethical considerations whether all the collected data is indeed of interest (Burchartz et al., 2020). Second, the definition of a valid day of accelerometer data via the specification of a minimum duration of wear time also has far-reaching implications. One study employed a semisimulation approach with data from the 2005-2006 NHANES, comparing a reference set with a wear time of 14 h/day to sets with 10, 11, 12, and 13 h/day (Herrmann et al., 2014). Significant differences were found in moderate physical activity between the reference set and the sets with 10 and 11 h/day wear time, but not for vigorous physical activity. Generally, with increasing wear time, participants were judged to take more steps, spend more time inactive, and in light or moderate intensity physical activity. Apart from the specific cut-off differentiating between a valid and an invalid day of accelerometer data, the way of determining whether this threshold was met or not is also relevant. A comparison of three different wear time algorithms found effects on not only the estimates of wear time, but also the associated sample size and estimates of physical activity indicators (Keadle et al., 2014).

However, this issue may be more pronounced for mail-based protocols, which the described study was, compared to situations where participants are personally handed the monitors by experimenters. Still, the question of trickle-down effects linked to the exclusion of some participants based on a given criterion of invalid data provided is highly relevant considering systematic differences may exist between people with valid and invalid data (Loprinzi et al., 2013), as this may impede generalisability. All of this variability in accelerometer study procedures, with no one standardised protocol, results in considerable difficulty in comparing studies utilising accelerometers (Burchartz et al., 2020; Pedišić & Bauman, 2015). Moreover, further adding to this issue is that fact that not all relevant decisions are necessarily exhaustively reported (Migueles et al., 2017; Montoye et al., 2018).

Despite the issues that may arise from such methodological heterogeneity, objective physical activity monitors still yield many unique advantages that deserve recognition. They can provide a very detailed record of physical activity (Ainsworth et al., 2015; Burchartz et al., 2020; Matthews, 2005), including frequency, duration, pattern, and intensity of activity (Ainsworth et al., 2015), which can be pinpointed to a specific time point (Strath et al., 2013). Some researchers have highlighted the limited invasiveness associated with accelerometry (Ainsworth et al., 2015); however, others have pointed to the burden placed on participants by requiring them to wear the devices for the duration of the study period (Silfee et al., 2018). Moreover, while a lack of error due to participants or researchers has been suggested before (Matthews, 2005), it has also been proposed more recently that it is actually possible for participants to more or less deliberately influence their physical activity estimates by not wearing the accelerometer for extended periods of time, modifying their actual behaviour, or even mishandling the device, e.g. by shaking it (Pedišić & Bauman, 2015).

The preceding section was intended to illustrate that physical activity is a complex behaviour. The various options for its measurement each afford the opportunity for unique insights as well as challenges, awareness of which is necessary for an informed engagement with the topic. Although objective measures are often considered more accurate than subjective measures, they should be deemed intricate tools which provide only approximations of actual physical activity behaviour (Burchartz et al., 2020). In particular, objective monitors may also be affected by measurement reactivity, which could lead to biased physical activity estimates. This thesis will now turn back to the topic of measurement reactivity, this time specifically in terms of what is known about measurement reactivity in physical activity assessed with objective activity monitors.

Measurement Reactivity in Physical Activity Measured With Wearable Monitors

Findings on Measurement Reactivity in Physical Activity

Some of the seminal research on measurement reactivity in physical activity utilised a within-person experimental design in which the true study purpose was initially masked. For instance, in one study participants were asked to wear a “body posture monitor” for 1 week, which was in fact a sealed pedometer with the daily step count display covered (Clemes et al., 2008). Then, after a debriefing in the laboratory, participants were requested to wear the pedometer for another week, this time unsealed, and to record their steps per day in an activity log. During Week 2, in the unsealed condition, mean steps per day were 11,385 (SD = 3,763), and thus significantly higher than the mean of 9,541 steps per day (SD = 3,186) during the first week. Within each condition, there were no differences in step counts between monitoring days. Since there were a number of proposed potential mechanisms for these findings, including participants’ awareness of being monitored with a pedometer in the 2nd week, being able to see the number of steps they took per day, as well as noting this information in the activity log, a follow-up study was conducted to allow for a better differentiation between these explanations. There, after a similar protocol as described above for the 1st week of the study period, termed the covert condition, participants completed three more conditions for 1 week each, namely another sealed condition, an unsealed condition without an activity log, and an unsealed condition in which daily steps were recorded in an activity log (Clemes & Parker, 2009). The order of the three conditions was balanced over participants. Mean steps per day were significantly higher in the activity log condition (M = 9,635; SD = 2,709) compared to both the covert (M = 8,362; SD = 2,600) and the sealed condition (M = 8,832; SD = 2,845) for the 2nd week. While there were no significant differences for mean steps per day between the sealed and unsealed conditions compared to the covert condition, there were differences between individual days in the former conditions, with a significantly higher number of steps on the 1st day compared to the 5th and 6th day in the sealed condition and the 5th and 7th day in the unsealed condition. As there were no effects of study day evident in the activity log condition, the authors hypothesised that measurement reactivity effects could be present for even longer than 1 week in conditions where the elicited effects are especially pronounced. Thus, a third in this series of studies examined the development of daily steps over 2 weeks of participants wearing unsealed pedometers and completing a log of daily steps after 1 week of covert monitoring like described before (Clemes & Deans, 2012). While mean daily steps were found to be significantly higher during the first week of the activity log condition (*M* = 9,898; *SD* = 3,002) compared to the covert condition (*M* = 8,331; *SD* = 3,010) and the second week of the activity log condition (*M* = 8,226; *SD* = 3,170), the latter two conditions did not significantly differ from each other. It was therefore concluded that measurement reactivity effects were unlikely to exceed 1 week.

More recently, building on a rapid review of measurement reactivity to objective measurement of health behaviour (French et al., 2021), a systematic review and meta-analysis examined measurement reactivity in digital in-the-moment measurements of health behaviours, including physical activity (König et al., 2022). Most of the reviewed studies on measurement reactivity in physical activity were observational studies (68%) examining the course of physical activity over a given study period without any explicit experimental manipulations. Linear trends of diminishing physical activity were taken as an indication of measurement reactivity in some studies (e.g. Baumann et al., 2018; Ullrich et al., 2021), while others compared physical activity between specific days of the study period with the expectation that Day 1 or Days 1 and 2 would be the most active (e.g. Craig et al., 2010; Dössegger et al., 2014). One subsequent study has examined both of these indicators simultaneously (Arigo & König, 2024). Of the 25 studies included in the systematic review, 10 utilised pedometers to measure physical activity and 15 accelerometers, with the more recent studies tending towards accelerometers (König et al., 2022). Whether evidence for measurement reactivity could be detected depended partly on the precise physical activity indicator being examined. Only two of nine studies assessing physical activity as moderate-to-vigorous physical activity reported significant findings. Out of six studies investigating measurement reactivity in light physical activity, three showed a significant effect. Measurement reactivity seemed to be the most consistently observed in daily steps, for which ten out of 15 studies in the systematic review found evidence. For activity counts, five out of the reviewed seven studies suggested measurement reactivity effects, which tended to be small, however, with Cohen's *d* ranging from 0.04 to 0.11. These latter two indicators of physical activity may thus be most prone to measurement reactivity, potentially because they are more easily modifiable than other indicators like moderate-to-vigorous physical activity, for example.

In order to determine overall effect sizes, a meta-analysis was conducted on seven studies experimentally manipulating measurement reactivity, resulting in an effect of Cohen‘s $d = 0.27$, 95% CI [0.16; 0.39] for a pooled estimate of all physical activity indicators, though the effects were quite heterogenous, with $I^2 = 78\%$ (König et al., 2022). Restricting the analysis to the four studies with daily steps as an indicator of physical activity yielded comparable results, with a slightly higher Cohen's $d = 0.30$, 95% CI [0.12; 0.49] and $I^2 = 86\%$. On the whole, it thus appears that measurement reactivity effects of up to medium size may indeed exist in physical activity. Considering that small effect sizes should be considered the norm in psychological research (Götz et al., 2022), this finding may be quite concerning. However, as has been illustrated not long ago, if measurement reactivity effects and especially their magnitude are well understood, it is possible to statistically control for them and consequently minimise their accompanying bias (Bendtsen & McCambridge, 2021). Still, what is as of yet relatively poorly understood are the mechanisms by which measurement reactivity operates.

Proposed Mechanisms of Measurement Reactivity

There have been a number of different means by which measurement reactivity effects have been speculated to function both in general and concerning physical activity in particular. One such mechanism concerns heightened self-monitoring (French & Sutton, 2010). The act of research participation thus may induce participants to be more aware of their own behaviour, potentially leading to the detection of an incongruity between actual and desired behaviour, which may be subsequently reduced (Arigo & König, 2024; König et al., 2022). As such, it has been suggested that measurement may operate like a behaviour change technique (French et al., 2021; French & Sutton, 2010), which are those elements of behavioural interventions responsible for enacting change (Michie et al., 2015). Indeed, one of the behaviour change techniques in a comprehensive taxonomy explicitly refers to self-monitoring as its core component (Michie et al., 2015). Still, this explanation by itself may not provide sufficient detail on what exactly the development of measurement reactivity entails.

Following König et al. (2025), a differentiation can be made between measurement reactivity operating via the introduction of researcher observation and the introduction of the measurement procedure. Consistent with the hypothesis that the observation of participants by members of the research team is central for the creation of measurement reactivity is the suggestion that social desirability and social norms are relevant in this context (Arigo & König, 2024; Baumann et al., 2018; Dössegger et al., 2014; König et al., 2022). Support for the potential implication of social desirability in measurement reactivity comes from findings that socially desirable or healthy behaviours may tend to be initially increased during a research study and socially undesirable behaviours may tend to be decreased (French et al., 2021). Such an account also closely mirrors what has been proposed as the core mechanism of the Hawthorne effect, though there are heterogenous explanations associated with this term (McCambridge, Witton, & Elbourne, 2014).

In terms of measurement reactivity associated with the measurement procedure itself, the novelty of being measured in general or of a specific measurement device may be a relevant factor for the creation of measurement reactivity effects (Baumann et al., 2018; Dössegger et al., 2014). Such an explanation is consistent with the finding that measurement reactivity may wear off after the first few days, as the novelty of measurement presumably fades (Arigo & König, 2024; French et al., 2021). Again, previous discussions surrounding the Hawthorne effect have also included mentions of novelty as a potential operating mechanism (McCambridge, Witton, & Elbourne, 2014).

The Present Study

The present study aims to further investigate these two mechanisms by building on the work of König et al. (2025). They recruited participants who were using personal fitness apps and trackers to monitor daily steps taken even before the start of the study and continued to do so during the study period. Participants completed two online surveys 2 weeks apart, filling in the number of steps recorded by their personal devices for the past 14 days at each date. In this manner, only researcher observation was introduced, but no novel measurement. No significant changes in steps per day could be detected between the 14 days before the commencement of the study and the 14 days of the study period, nor Day 1 of the study period or Day 1 and 2 taken together compared to the other days under investigation. It was thus concluded that no meaningful measurement reactivity exists in people with experience with physical activity trackers using their own devices. However, no direct comparison was conducted between people with and without prior physical activity tracking experience. Such a contrast could further elucidate the contribution that prior physical activity tracking experience has on measurement reactivity since presumably the act of monitoring one's physical activity has less novelty value for experienced physical activity trackers.

Concerning the other potential mechanism, namely researcher observation, the study by König et al. (2025) also included a randomised allocation of participants to either a high salience of researcher observation condition, in which the fact that physical activity would be observed was emphasised, or alternatively a condition with low salience of researcher observation without this information. No differences were found between steps per day in the two salience conditions. However, it was speculated that the salience manipulation may not have been very powerful as it only occurred in an online survey with no face-to-face contact with researchers. It was therefore worth investigating whether a salience manipulation delivered in person by a member of the research team might be more impactful and capable of producing measurement reactivity.

In short, the aim of the present study was to further the research by König et al. (2025) by investigating the two proposed mechanisms for measurement reactivity in physical activity: the novelty of measurement, through a direct contrast of people with and without prior experience with physical activity tracking, for whom this novelty presumably differs, and the salience of researcher observation, intended to be more powerful for being delivered in person. Gaining a better understanding of the inner workings of measurement reactivity could enable a better comprehension of the precise circumstances in which measurement reactivity might be problematic and in designing studies in a way so as to minimise bias when necessary.

To this end, the following research questions and hypotheses were investigated:

Research Question 1

Does measurement reactivity in physical activity differ between novice and experienced physical activity trackers?

Hypothesis 1: Novice trackers will show a steeper decline in physical activity over the study period compared to experienced trackers.

Research Question 2

Does salience of researcher observation affect measurement reactivity in physical activity in novice and experienced physical activity trackers?

Hypothesis 2: High salience of researcher observation will result in a steeper decline in physical activity over the study period compared to low salience.

Research Question 3

Does salience of researcher observation affect measurement reactivity in physical activity in novice and experienced physical activity trackers differently?

Hypothesis 3: The effect of high salience of researcher observation resulting in a steeper decline of physical activity over the study period will be more pronounced in novice compared to experienced trackers.

Methods

This thesis was preregistered in the health psychology research group's internal study database before data analysis was conducted. Materials, data, and analysis scripts for this study are documented in the health psychology research group's study database. This study forms part of a larger research project (König et al., 2024), which also involves two other master's theses on measurement reactivity in physical activity focussing on measurement reactivity in self-reported physical activity (Schmikat, 2025) and specifically in experienced physical activity trackers, respectively (Batek, 2025). The ethics committee of the University of Vienna approved the study protocols for the larger research project. In the following sections, descriptions of methods are mainly limited to those aspects of the research project that are relevant for the present study.

Participants

Recruitment for this study targeted two groups of people: experienced users of fitness apps and trackers who had routinely recorded their daily steps for a minimum of 14 days prior to enrolment and novice trackers who had never used any kind of fitness app or tracker before. Although it is possible to examine levels of physical activity tracking experience with more nuance, for example by considering different stages of the adoption process of such trackers (König et al., 2018), for this study a stricter definition and consequent recruitment of people on the two ends of the experience spectrum were utilised to generate more easily interpretable effects. Therefore, people were excluded from participation if they reported having tracked their physical activity occasionally in the past, but were not currently doing so at the start of the study. Participants also had to be aged 18 years or over and have sufficient German language skills to understand the study instructions and questionnaires. Furthermore, participants were excluded if they reported any major illnesses preventing them from engaging in physical activity in the 14 days prior to enrolment.

Participants were mainly recruited via the University of Vienna Laboratory Administration for Behavioural Sciences (LABS) system, which connects students who gain credits for participating in studies as part of their degree with researchers based at the university looking for study participants. Additionally, some participants were recruited via word-of-mouth. Participants recruited via LABS received 6 credits upon completion of the study. Participants recruited elsewhere had the opportunity to enter a lottery to win one of three vouchers of 100€ for wunschgutschein.at. During university holidays, no students were enrolled in the study since it was presumed that patterns of physical activity in these periods could differ from everyday activity during the semester.

No explicit power analysis was conducted for this study, in part because power analyses for multilevel modelling are very complicated due to the number of parameters involved (Hoffman, 2015; Snijders & Bosker, 2012). Since the comparison of novice and experienced physical activity trackers is a novel paradigm of investigating measurement reactivity in physical activity, reasonable estimations for these parameters are difficult. Although the impact of researcher salience on measurement reactivity has been examined before, it is still questionable whether the reported results can be transferred to the present study, which differs from the previous study in several aspects. For the larger research project which this study is a part of the aim is to recruit 330 participants. This calculation was based on a repeated measures ANOVA of 6 groups and 2 measurements with an effect size of $f(V) = 0.2$ with $\alpha = 0.05$ and $1\text{-}\beta = 0.8$, conducted in G*Power. As the larger project is ongoing at the time of writing, participants who completed the study from November 2024 until April 2025 were included in the present analyses. Only the subgroups relevant to the present study were utilised.

Design and Procedure

This study followed a longitudinal 2 (tracking experience: novice/experienced tracker) x 2 (salience: high/low) between-subject design. It was a mixture of a true experiment (salience) with a natural experiment (tracking experience). It should be noted that the larger research project which the present investigation is a part of includes a further group differentiation based on whether experienced trackers received a research-grade accelerometer or continued using solely their own physical activity tracking devices. The present study only utilised data from those experienced trackers who also received accelerometers and could thus provide accelerometer-based daily step counts.

The study consisted of two session in the laboratory exactly 14 days apart. Sessions were conducted with one participant at a time. The first session started with a brief description of the masked study purpose, indicating that study aim was merely about physical activity in everyday life. Participants were thus not made aware that the study was about measurement reactivity. Next, participants were asked about their prior experience with physical activity tracking, with the experimenter determining whether they qualified as experienced or novice trackers according to the criteria described above. Then, after age and lack of illnesses impeding physical activity in the past 14 days had been confirmed by the participants, the provision of informed consent followed. Participants then created their personal code for pseudonymisation. Next, they filled out an online questionnaire including demographic information, prior physical activity tracking experience, and questions about injuries, illnesses or other events that prevented participants from engaging in their usual level of physical activity in the past 14 days. Included in this survey were also questions about physical activity in the past 14 days, which were, however, not utilised in the present study. During the completion of the survey, participants were randomly assigned to a salience condition. To this end, a sealed envelope was drawn from a stack of randomly shuffled envelopes containing information on salience condition. Block randomisation was utilised to ensure a roughly equal number of participants for each salience condition in every block of 45 participants. Afterwards, participants received one of four different study information sheets, depending on the assigned condition (high/low salience and experienced/novice tracker), which the experimenter read through with each participant. This sheet contained the salience manipulation, information on the second session in the lab, what to do until then, and how to use the accelerometer. People with physical activity tracking experience were asked to continue using their own tracking devices. Lastly, weight and height were measured by the experimenter before the physical activity sensor was activated.

The second session also included an online survey with questions about injuries, illnesses or other events that prevented participants from engaging in their usual level of physical activity in the past 14 days. Similarly to the first session, participants were also asked to report on their physical activity during the past 14 days, but this information was not used in the current analyses. Then followed a full debriefing about the purpose of the study, namely the investigation of measurement reactivity, and a brief description of the hypotheses under investigation.

Measures and Materials

Prior Physical Activity Tracking Experience

The key independent variable for Hypotheses 1 and 3 was prior physical activity tracking experience, which was self-reported by participants in the first session. Participants were categorised either as experienced physical activity trackers or novice physical activity trackers. Tracking experience was thus operationalised as a dichotomous variable. Although participants were not explicitly made aware of which group they were assigned to, it was specified during the recruitment process that specifically experienced and novice trackers were sought. Consequently, there was no strict masking of condition for tracking experience. The experimenter was also aware of a participant's tracking experience categorisation.

Salience of Researcher Observation

For Hypotheses 2 and 3, the central independent variable was salience of researcher observation. Participants were randomly assigned to either the high or low salience condition. Salience condition was masked for participants, but the experimenter was aware of condition assignment, since they read out the salience manipulation and handed the participant the appropriate study information sheet. The text in the information sheet for each salience condition can be found in Table A1 in Appendix A.

Study Day

Another fundamental independent variable for all hypotheses was the day in the study period. This corresponds to the time variable in the longitudinal model. Study day was based on the date recorded by the accelerometer's internal clock. For more details on the data processing, see Statistical Analyses.

Steps per Day as an Accelerometer-Derived Physical Activity Indicator

The key dependent variable for all hypotheses was steps per day. Steps per day was chosen as an indicator of physical activity since it is likely to be impacted by measurement reactivity, and thus well fitted for explorations of specific mechanisms of this phenomenon. If a reactivity effect is unlikely to occur to begin with, or only very small, manipulations of reactivity are likewise improbable to introduce a discernible effect. The decision was made to utilise accelerometer-based step counts for both tracking experience groups for optimal comparability. Steps per day were determined by movisens Move 4 accelerometer. Data processing was conducted with the movisens DataAnalyzer software using the Step Count algorithm. Move 4 has demonstrated high inter-instrument reliability with other accelerometers commonly used in physical activity research (ActiGraph GT3X+, GENEActiv, and Axivity AX3) when processing of raw data is done in an identical manner (Migueles et al., 2022). The wear time protocol was limited to waking hours since only physical activity was a variable of interest and not sleep, so a 24-hour protocol could be considered unethical (Burchartz et al., 2020). As recommended by Montoye et al. (2018), key accelerometer data collection and processing information is summarised in Table 1.

Table 1

Accelerometer Data Collection and Processing Specifications

Reporting item	Specifications for this study
Brand and model of the accelerometer	movisens Move 4
Epoch length for data collection and analysis	data collection: 64 Hz
Placement of accelerometer and side of body	right hip
Number of participants receiving accelerometer	48 participants
Accelerometer distribution method	in person
Days of data collected at each time point	14 days, all waking hours
Criteria for defining non-wear of accelerometer	determined by the movisens DataAnalyzer software's NonWearTime algorithm
Number of valid days and number of minutes per day of accelerometer data needed to be included in analysis	≥ 10 hours per day and ≥ 8 days
Accelerometer data PA outcome of interest and the interpretation method	steps per day as determined by the movisens DataAnalyzer software's StepCount algorithm
Number of participants non-compliant or who had accelerometer malfunction issues	Fifteen participants were excluded from the analysis because they did not meet the wear time criterion; two participants dropped out of the study and their data was thus also excluded.

Covariate Weekend/-day

For all hypotheses a covariate of weekend/-day was included. Physical activity indicators including steps have repeatedly been found to differ in adults depending on the day of the week (Clemes & Parker, 2009; Klenk et al., 2019; McVeigh et al., 2016; Tudor-Locke et al., 2004). Though some heterogeneity exists concerning which days precisely are the least active and the most dissimilar from other days, for this study a dichotomous contrast between weekdays (Monday through Friday) with weekends (Saturday and Sunday) was decided upon.

Sociodemographic Variables and Sample Description

The following sociodemographic variables were utilised in the present study: age, gender, highest education, student status, household net income. Moreover, BMI as calculated from weight and height measured by the experimenter was also included in the sample description. Furthermore, the degree of experience of experienced trackers with their personal physical activity tracking devices was quantified with duration of tracker use and self-reported consistency of tracker use on a scale of 1 (*not at all consistent*) to 5 (*very consistent*).

Statistical Analyses

All data analyses were conducted in R 4.5.0. For all inferential statistical testing, α was set to .05.

Data Cleaning and Preprocessing

Data cleaning involved a harmonisation of participant's pseudonymisation codes recorded in the online questionnaire and noted by the experimenter in the accelerometer data file. Differences in capitalisation were resolved and any German Umlauts were substituted. Moreover, one participant submitted two different durations for the length of their physical activity tracker use because they used two devices; the longer duration was included for the current analyses. The necessity of these data cleaning procedures had not been anticipated and thus these steps were not preregistered.

Extrapolating from guidelines by Migueles et al. (2017), data was excluded from participants who failed to provide valid accelerometer data for ≥ 10 hours per day and ≥ 8 days over the 2 weeks. Accelerometer wear time was calculated by taking the accelerometer-determined record duration for each day and subtracting the non-wear time detected by the Move 4 accelerometer and processed according to the movisens DataAnalyzer software's NonWearTime algorithm. Additionally, based on measured participant weight (in kg) and height (in cm), each person's BMI was calculated as kg/m^2.

Sample Description and Randomisation Checks

The following descriptive statistics were calculated for the whole sample and separately for novice and experienced physical activity trackers, and for the high and low salience groups: absolute number and relative frequency of participants in each tracking group and salience condition, age, gender, highest level of education, household net income, whether or not they were a student, and BMI. Mean accelerometer wear time and steps were also summarised descriptively for each day and the sum of all days. Additionally, presence of respiratory illness, injury, or stressful life events during the study period as relayed by participants in the second session were documented. For experienced physical activity trackers, the duration of physical activity tracker usage in months as well as self-reported consistency of tracker usage in general were also reported. In the preregistration, the separate reporting of the descriptive variables was only explicitly specified for the two tracking experience groups and not the salience conditions; however, since it was predetermined that randomisation checks utilising these same variables should be conducted, their full detailing by salience condition was also implied. Randomisation checks were conducted both for the two tracking experience groups and the two salience conditions using these descriptive variables. To examine differences in age, BMI, and accelerometer wear time, 2 x 2 analyses of variance with tracking experience and salience condition as independent variables were conducted. This constitutes a deviation from the preregistration, where it was suggested that t-tests be conducted for all metric variables. T-tests were utilised only to examine differences in the duration of physical activity tracker usage in months and the self-reported consistency of tracker usage in general between experienced trackers in the two salience conditions. To examine differences in gender, highest level of education, student status, household net income, presence of respiratory disease, injury, or stressful life events in the past 14 days, Chi^2-tests were conducted. Lastly, the intraclass correlation for steps per day, describing the proportion of between-person variation versus within-person variation of steps per day over the study period was also calculated.

Inferential Statistics

For all hypotheses, multilevel models were utilised, with steps per day on Level 1 nested in participants on Level 2. All analyses included the covariate of weekday versus weekend, with Saturday and Sunday = 1 versus all other days of the week, i.e. Monday to Friday = 0. Maximum likelihood estimations were used for all models. Calculations relied on the R packages lme4 (Bates et al., 2025) and lmerTest (Kuznetsova et al., 2022). In a deviation from the preregistration, the last day of the study period was excluded for the main models reported in the Results section. Only very little wear time and steps were provided by participants on this day compared to the other days (for descriptive wear time and steps on each day, see also Table 3 for the full sample as well as Table A 3 and Table A 4 in Appendix A for experience groups and salience conditions separately) as this constituted only half a day before the accelerometers were handed back to the study team in the second session of the study. Because an unconditional model of change did not converge with random slopes for time, i.e. study day, the subsequent models of interest were estimated without random slopes.

The only random effects included pertained to random intercepts. As a sensitivity analysis, models including the last day were also calculated and are reported in Table A 5, Table A 6, and Table A 7 in Appendix A. In a further deviation from the preregistration, the models of interest for hypothesis-testing were additionally examined excluding the first day of the study period as well as the last day as a sensitivity analysis since this, too, was not a full day. The results for these models are included in Appendix A in Table A 8, Table A 9, and Table A 10. For both sets of models in the sensitivity analysis, the unconditional models with only time as a Level-1 predictor did not fit significantly better with random slopes for the effect of time than without, so no random slopes were estimated for the subsequent models of interest for the sake of model parsimony. Thus, only random intercepts were included. In a further deviation from the preregistration, simple slope analyses using the R package interactions (Long, 2025) were conducted to better disentangle the interaction effects in the main models.

Assumption Checks. Assumptions for the models utilised for hypothesis-testing include the following (Shaw & Flake, n.d.): The functional form of the relationship between outcome and predictor(s) is linear. Level-1 residuals are independent and normally distributed. Level-2 residuals are independent and follow a multivariate normal distribution. Level-1 and Level-2 residuals are unrelated. Predictors on either level are not related to residuals on the other level. Assumption checks were conducted as recommended by Shaw and Flake (n.d.), with assumptions concerning linearity, independence of residuals and homoscedasticity being inspected with scatterplots and tests for correlation; and assumptions concerning (multivariate) normality being inspected with histograms and Q-Q plots.

Hypothesis 1. Hypothesis 1 predicted that novice trackers would show a steeper decline in physical activity over the study period compared to experienced trackers. To test this hypothesis, a multilevel model was specified with steps per day as the outcome. On Level 1, a linear effect of study day (day 1-14, coded as day 1 = 0 through day 14 = 13) was included as a predictor of time. Prior physical activity tracking experience (no = 0/yes = 1) was included as a predictor on Level 2. The key model parameter of interest was a cross-level interaction of study day by prior physical activity tracking experience.

Hypothesis 2. Hypothesis 2 predicted that high salience of researcher observation would result in a steeper decline in physical activity over the study period compared to low salience. To test this hypothesis, a multilevel model was specified with steps per day as the outcome. On Level 1, a linear effect of study day (day 1-14, coded as day 1 = 0 through day 14 = 13) was included as a predictor of time. Salience condition (low = 0/high = 1) was included as a predictor on Level 2. The key model parameter of interest was a cross-level interaction of study day by salience condition.

Hypothesis 3. Hypothesis 3 predicted that the effect of high salience of researcher observation resulting in a steeper decline of physical activity over the study period would be more pronounced in novice compared to experienced trackers. To test this hypothesis, a multilevel model was specified with steps per day as the outcome. On Level 1, a linear effect of study day (day 1-14, coded as day 1 = 0 through day 14 = 13) was included as a predictor of time. Both prior physical activity tracking experience (no = 0/yes = 1) and salience condition (low = 0/high = 1) were included as predictors on Level 2. The key model parameter of interest was a three-way interaction of study day by salience condition by prior physical activity tracking experience.

Effect Sizes for Hypotheses 1 to 3. Effect size calculations were based on the framework by Rights and Sterba (2019) and their extension for longitudinal data (Rights & Sterba, 2021). Based on the basic premise that R^2 measures reflect the proportion of outcome variance that can be explained by the model, a full decomposition of R^2s is calculated based on the different potential sources of explained variance and the outcome variance of interest. The outcome variance may refer to either total outcome variance, within-person variance, or between-person outcome variance. A subscript *t*, *w*, and *b*, respectively, represents these outcome variances in the shorthand notation employed subsequently. Variance may be explained by Level-1 predictors via the fixed components of slopes (f_1), Level-2 predictors via the fixed components of slopes (f_2), Level-1 predictors via random slope variation (v), and person-specific outcome means via random intercept variation (m). These sources to which variance may be attributable are indicated in the superscript in the shorthand notation employed subsequently.

The definition and interpretation of the sources of explained variance explained here were derived under the assumption of person-mean-centring. When employing other centring strategies for Level-1 predictors including time, namely some form of centring-at-a-constant, definitions can differ. However, if time and other Level-1 predictors are balanced and thus do not differ between-person, then the somewhat simpler definitions from person-mean-centring can be retained. In the present analyses, two Level-1 predictors are included, neither of which are person-mean-centred: study day, centred at origin, and the weekend/weekday covariate, dummy-coded so that 0 reflects a weekday. Still, in the present study, time, i.e. study day, and the covariate are nearly balanced over participants. Out of 434 observations from 31 participants over Days 1-14 (excluding the last day), 1 observation for daily steps was missing, less than 1%. The estimated variance attributable to f_2 in an unconditional model of change including only study day and the weekend/weekday covariate was 0 (see also Table A 15 in Appendix A). As there was virtually no between-person variability in the included Level-1 predictors, the definitions of f_1 and f_2 from person-mean centring could thus be retained.

As preregistered, the proposed set of R^2 measures were computed for each of the three models specified for Hypotheses 1-3. Additionally, a series of model comparisons via ΔR^2s were conducted to investigate the unique contributions of the added predictors, first going from an unconditional model of change with only study day and the weekend/-day covariate as Level-1 predictors to a conditional model with one Level-2 predictor, separately with either prior physical activity tracking experience or salience, to quantify the variance explained by the Level-2 predictor and its interaction with time. Then, in a slight deviation from the preregistration, a reduced model differing from the final model for Hypothesis 3 only in the lack of the three-way interaction between study day, experience, and salience, was computed and compared to the final model for Hypothesis 3 to quantify the variance explained specifically by the three-way interaction. Calculations were done with the R package r2mlm (Shaw et al., 2025).

Results

Descriptive Statistics

Out of 46 participants who completed the study and were eligible for the present analyses since they received accelerometers, data from 31 participants who met the specified wear time criteria were included. Participants were aged between 18 and 57 years, with an average of 22.87 years (*SD* = 7.62). Twenty-three participants (74.19%) identified themselves as female, eight as male (25.81%), and none as another gender. Selected descriptive sample characteristics including demographic variables and BMI as well as presence of respiratory illness, injury, or stressful life events during the study period as reported by participants in the second session are presented in Table 2 for the whole sample and in Table A 2 in Appendix A separately according to tracking experience and salience condition. Additionally, mean tracker use and consistency of tracker use are also reported for experienced tracker’s own devices for all experienced trackers in Table 2 and by salience condition in Table A 2 in Appendix A. Steps per day as the key outcome together with accelerometer wear time are summarised descriptively in Table 3 for the whole sample. Table A 3 and Table A 4 in Appendix A show accelerometer wear time and daily steps, respectively, for experienced and novice physical activity trackers as well as the high and low salience conditions separately. Figure 1, Figure 2, and Figure 3 moreover illustrate the progression of steps per day for the whole sample, and separately by tracking experience and salience condition, respectively.

Randomisation checks tested for differences between the two tracking experience groups and the two salience conditions, respectively. Neither the tracking experience groups nor the salience conditions differed in gender, highest level of education, student status, household net income, or presence of respiratory illness, injury, or stressful life events during the study period, as can be seen in Table A 2 in Appendix A. Concerning age, there was a significant effect of tracking experience, but not of salience (see also Table A 2), and no significant interaction ($F(1,27) = 2.3$, $p = .41$). Since physical activity is usually inversely associated with age (Bauman et al., 2012), but in this sample, the older group were experienced trackers rather than novice trackers, who tended to perform more physical activity in terms of steps taken each day both descriptively (see also Table A 4) and in the models for hypothesis testing (see also Table 4 and Table 6), it was not deemed necessary to adjust the models for age. For BMI, there was no significant effect for either tracking experience or salience (see also Table A 2), but there was a significant interaction ($F(1,27) = 4.3$, $p = .048$). However, Tukey's HSD post hoc tests did not yield significant differences between any two groups. As evident in Table A 3 in Appendix A, no differences in accelerometer wear time were detected between either tracking experience groups or salience conditions over the study period with the only exception of Day 9, where significant differences between salience conditions (see also Table A 3) and a significant interaction between experience and salience ($F(1,27) = 6.49$, $p = .017$) could be attested.

The intraclass correlation (ICC) was calculated for steps per day, with an ICC = .25 suggesting that relatively little overall variation in steps per day over the study period was attributable to between-person variation, in comparison to the proportion of the variance attributable to within-person difference.

Table 2

Descriptive Characteristics of Participants for the Full Sample

Descriptive variable	*n/M*	%/*SD*
Age	22.87	7.62
Gender		
female	23	74.19%
male	8	25.81%
other	0	0%
Highest level of education		
no school-leaving certificate	0	0%
Some high school, to 8th grade	1	3.23%
Some high school, no university entrance diploma	0	0%
High school graduate, university entrance diploma or the equivalent (Abitur/ Matura)	25	80.65%
Some college, no degree	3	9.68%
Trade/technical/vocational training	1	3.23%
Bachelor's degree	1	3.23%
Master's degree	0	0%
Doctoral degree	0	0%
Student [a]	29	93.55%
Household net income		
less than 150€	5	16.13%
150€ to 300€	2	6.45%
300€ to 500€	3	9.68%
500€ to 1,000€	8	25.81%
1,000€ to 1,500€	2	6.45%
1,500€ to 2,000€	2	6.45%
2,000€ to 2,500€	1	3.23%
2,500€ to 3,000€	5	16.13%
3,000€ to 5,000€	3	9.68%
5,000€ to 10,000€	0	0%
more than 10,000€	0	0%
BMI	23.49	2.97

Descriptive variable	*n/M*	%/*SD*
presence of respiratory disease in the past 14 days [a]	1	3.23%
presence of injury in the past 14 days [a]	2	6.45%
presence of stressful life events in the past 14 days [a]	11	35.48%
For experienced physical activity trackers:		
Duration of tracker usage in months	33.58	12.92
Self-reported consistency of tracker usage in general (1 = *not at all consistent* to 5 = *very consistent*)	4.08	1.16

Note. T-tests were conducted with Welch's correction.

[a] Absolute and relative frequencies of participants answering "yes".

Table 3

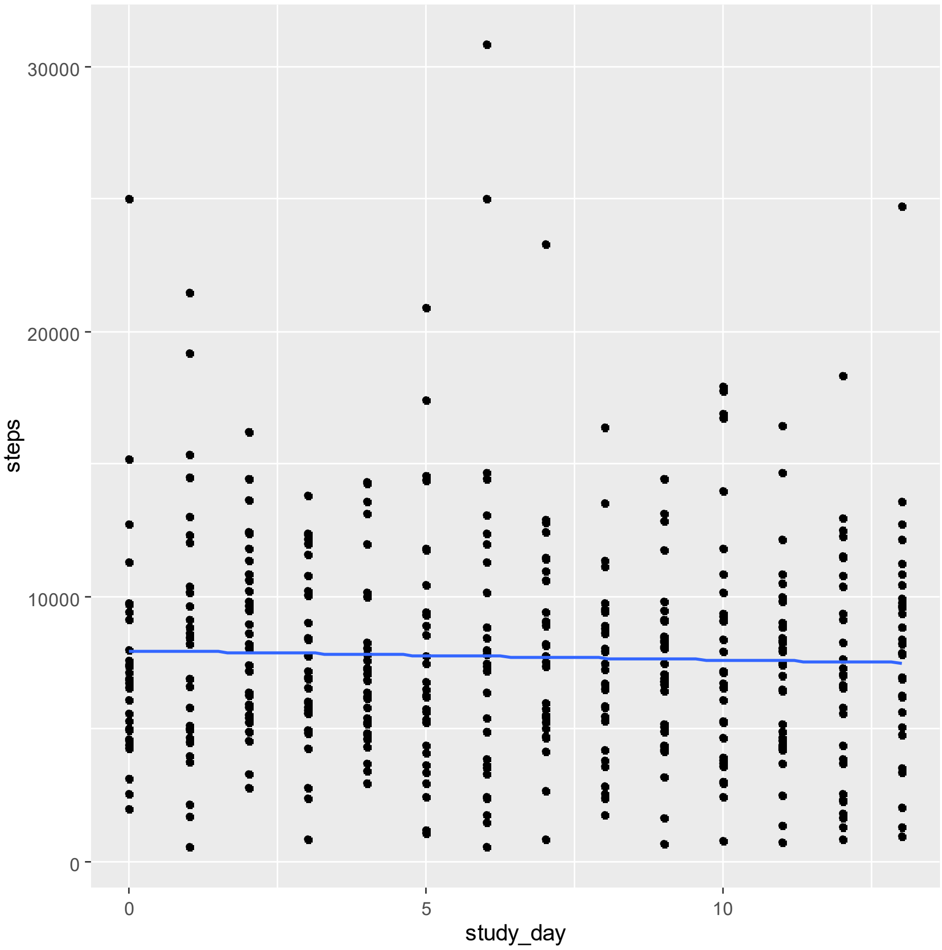
Accelerometer Wear Time and Steps per Day for the Full Sample

Day in study period	Accelerometer wear time		Steps per day	
	M	*SD*	*M*	*SD*
Over days 1-15	11:42:21	04:03:38	7,455	4,271
Over days 1-14	12:20:08	03:19:33	7,735	4,214
Day 1	09:53:50	02:14:22	7,409	4,369
Day 2	13:32:25	02:43:35	8,367	4,862
Day 3	12:31:58	02:49:11	8,667	3,347
Day 4	11:40:33	03:01:30	7,562	3,248
Day 5	12:37:50	03:30:35	7,461	3,229
Day 6	12:27:06	02:40:54	7,690	4,614
Day 7	12:48:15	03:40:04	8,026	6,749
Day 8	13:15:54	02:52:14	8,205	4,111
Day 9	12:00:31	03:06:29	7,091	3,399
Day 10	13:02:50	03:58:06	7,276	3,189
Day 11	12:25:21	03:50:59	7,980	4,716
Day 12	12:21:23	02:59:41	7,196	3,570
Day 13	11:31:58	04:19:16	7,149	4,183
Day 14	12:31:50	03:12:42	8,200	4,442
Day 15	02:53:31	03:03:34	3,273	2,649

Note. Accelerometer wear time in hh:mm:ss format.

Figure 1

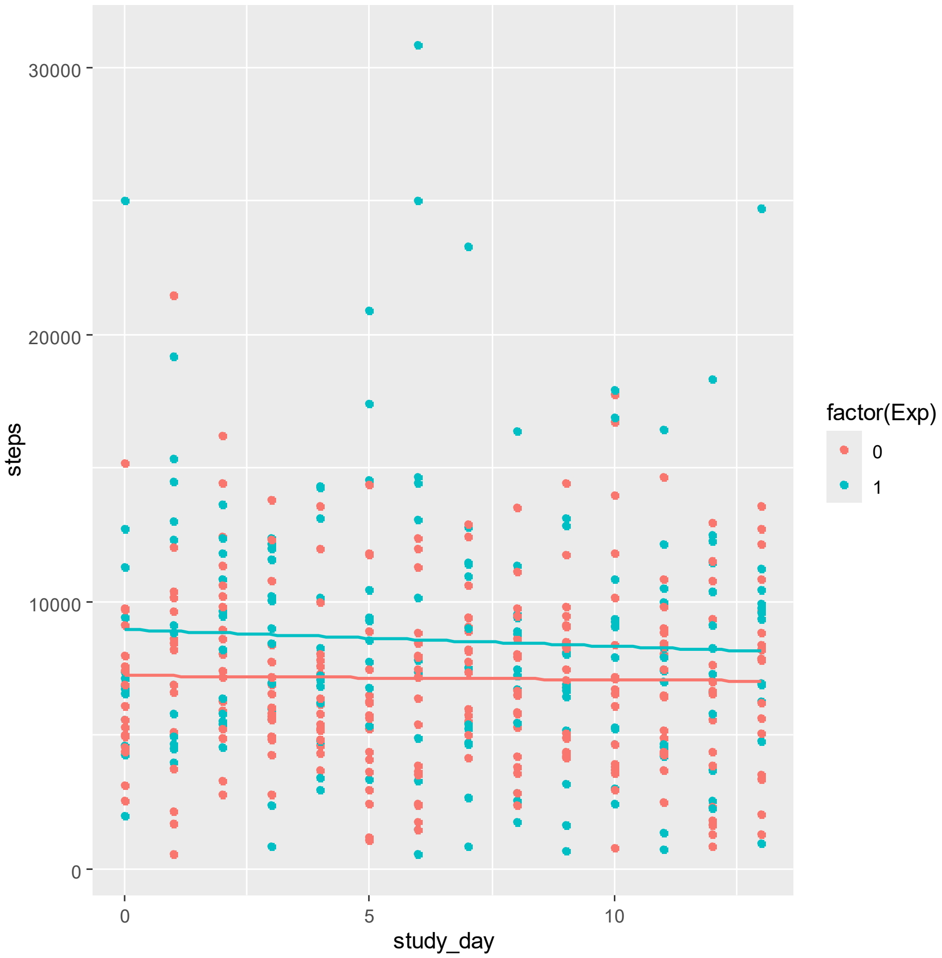
Steps per Day Over the Study Period for the Whole Sample



Note. The scatterplot shows steps per day on each day of the study period, with the line indicating the averaged progression over time. Study day 0 indicates the first day of the study period.

Figure 2

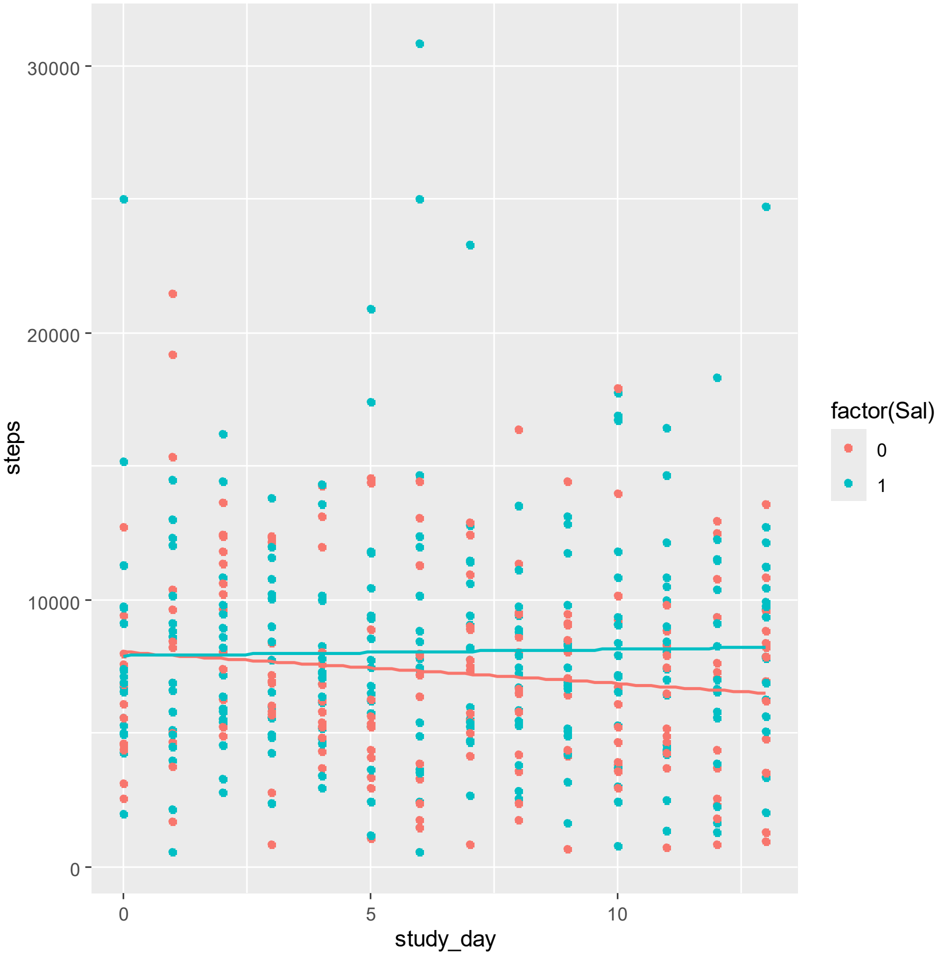
Steps per Day Over the Study Period Separately for Experienced and Novice Trackers



Note. The scatterplot shows steps per day on each day of the study period, with the lines indicating the averaged progression over time per group. Novice trackers are represented by orange dots and the orange line for Exp = 0, while experienced trackers are represented by blue dots and the blue line for Exp = 1. Study day 0 indicates the first day of the study period.

Figure 3

Steps per Day Over the Study Period Separately for Low and High Salience Conditions



Note. The scatterplot shows steps per day on each day of the study period, with the lines indicating the averaged progression over time per group. Participants in the low salience condition are represented by orange dots and the orange line for Sal = 0, while participants in the high salience condition are represented by blue dots and the blue line for Sal = 1. Study day 0 indicates the first day of the study period.

Inferential Statistics

Model Assumptions Checks

Assumption checks were conducted following recommendations from Shaw and Flake (n.d.) for the each of the three models of interest. No quadratic or other polynomial trends nor any points of inflection indicating inappropriateness of a linear trend were identified. For all models, independence of residuals, homoscedasticity, and (multivariate) normality could approximately be assumed for the most part. There were significant correlations between level-2 intercept residuals and level-1 residuals, which not deemed critical since they were relatively small with $r = .12$ in all models. It was thus concluded that the assumptions could broadly be considered fulfilled, indicating the appropriateness of employing the constructed models.

Hypothesis 1: The Effect of Experience on Steps per Day

Model Estimations. The model parameter estimates for the final model testing Hypothesis 1 can be found in Table 4. There was no significant effect of tracking experience, time and, mostly importantly for the hypothesis at hand, the interaction of time by tracking experience. Only the covariate comparing weekend to weekday achieved significance, with an average of 1,677 less steps taken on a Saturday or Sunday compared to Monday to Friday. The simple slope analysis indicated that the effect of time was not significant in either the novice trackers, for which estimates are reflected in the main effect of time in Table 4, or in experienced trackers ($b = -70.45$, $SE = 65.52$, $t = -1.08$, $p = .28$). Figure A 1 in Appendix A shows a Johnson-Neyman plot illustrating the interaction, or lack thereof, between time and tracking experience. The first hypothesis was thus rejected.

Compared to this model, a sensitivity analysis with a model including the last day contained a nearly significant simple main effect of time with $p = .052$ at an estimated decline of 102 steps each day in novice trackers. The main effect of tracking experience was moreover significant in the model including the last day, with experienced trackers taking an average of 1,980 steps more on Day 1 (see also Table A 5 for the full model parameters). Differences of the model presented here to the model without the first and the last day lie only in parameter estimates without additional significance of effects (see also Table A 8 for the full model parameters). Neither of the models within the sensitivity analysis indicated a significant interaction of time by tracking experience.

Effect Size Decomposition. In terms of the within-person outcome variance in the model for Hypothesis 1, $R_w^{2\,(f_1)}$ indicated that 4.54% were attributable to the Level-1 predictors, i.e. time and the weekend/-day covariate, as well as the study day by experience cross-level interaction, via fixed components of slopes. Concerning between-person outcome variance, $R_b^{2\,(f_2)}$ suggested that 10.91% may be attributable to the fixed effect of the Level-2 predictor prior physical activity tracking experience. According to $R_t^{2\,(f_1)}$, 3.41% of the total outcome variance were attributable to the Level-1 predictors study day and the weekend/-day covariate as well as the cross-level interaction of study day and tracking experience via fixed components of slopes. The proportion of total outcome variance attributable to the fixed effect of tracking experience, $R_t^{2\,(f_2)}$, was 2.72%. The proportion of total outcome variance attributable to all Level-1 and Level-2 predictors and their interaction via the fixed components of slopes, $R_t^{2\,(f)}$, was thus 6.13%. Figure A 4 in Appendix A serves as an illustration of the variance attributable to different sources in the model. The full set of R^2 measures, including those not mentioned here, can be found in Table A 11 in Appendix A.

Table 4

Model Effects for Hypothesis 1 (Excluding Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	7,689.56	640.8	70.71	12	< .001*
Time	-9.58	55.73	402.03	-0.17	.864
Tracking experience	1,800.48	977.13	67.44	1.84	.07
Interaction of Time * Tracking Experience	-60.87	86.05	402.02	-0.71	.48
Weekend/-day	-1,677.01	378.94	402.03	-4.43	< .001*
Covariate					
Random effects					
Intercept	3,934,803	1,984			
Residual variance	12,688,506	3,562			
Number of Parameters	7				
-2LL	8,362.9				
AIC	8,376.9				
BIC	8,405.4				

* p-values are significant at α = .05.

Hypothesis 2: The Effect of Salience on Steps per Day

Model Estimations. The model parameter estimates for the final model testing Hypothesis 2 can be found in Table 5. Again, the covariate comparing weekend to weekday achieved significance, with an average of roughly 1,634 less steps taken on a weekend day compared to a weekday. However, there was no significant effect of time, salience and no significant interaction of salience by time. The simple slope analysis suggested that there was no significant effect of time for either the low salience condition, as shown in Table 5, or the high salience condition (b = 14.02; SE = 55.66; t = 0.25, p = .80). Figure A 2 in Appendix A serves as a graphical illustration of the interaction, or lack thereof, between time and salience. Consequently, Hypothesis 2 was also rejected.

In comparison, the model including all days contained a significant effect of time at p = .002 with an estimated decrease of 193 steps per day in the low salience condition. However, there was also no significant main effect of salience and, more importantly, no interaction between time and salience (see Table A 6 for the full model parameters). In the model excluding both the first and the last day, the effect of time was also significant at p = .008, with an estimated fall in steps of roughly 190 per day in the low salience condition. Moreover, this model included a significant interaction between time and salience (p = .037), indicating a growing difference of 196 steps per day in the study period between the salience conditions, with the high salience group taking increasingly more steps compared to the low salience group. The simple main effect of time would thus be an increase of roughly 7 steps per day in the high salience group. However, the simple main effect of salience was not significant, indicating an insignificant group difference of the high salience group being less active by roughly 495 steps on Day 1. Although this model included a significant time by salience interaction, its direction is opposite to that proposed in Hypothesis 2. Full model parameters for the model excluding both the first and the last day are reported in Table A 9.

Effect Size Decomposition. In the model for Hypothesis 2, the proportion of within-person outcome variance attributable to the Level-1 predictors study day and the weekend/-day covariate and the Level-1 by Level-2 cross-level interaction of study day and salience via fixed components of slopes, $R_w^{2\,(f_1)}$, was 4.84%. $R_b^{2\,(f_2)}$ suggests that 3.43% of the between-person outcome variance were attributable to the fixed effect of the Level-2 predictor salience. As indicated by $R_t^{2\,(f_1)}$, 3.63% of the total outcome variance may be accounted for by the Level-1 predictors study day and the weekend/-day covariate and the cross-level interaction of study day and salience via the fixed components of slopes. The proportion of total outcome variance attributable to salience via the fixed components of slopes, $R_t^{2\,(f_2)}$, was 0.86%. As per $R_t^{2\,(f)}$, the proportion of total outcome variance attributable to all Level-1 and Level-2 predictors and their interaction via the fixed components of slopes was therefore 4.49%. Variance attributable to various sources in the Hypothesis 2 model are depicted in Figure A 5 in Appendix A. The full set of R^2s are described in Table A 12 in Appendix A.

Table 5

Model Effects for Hypothesis 2 (Excluding Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	8,418.15	765.45	66.12	11	< .001*
Time	-103.26	65.47	402.03	-1.58	.116
Salience	25.35	999.19	64.79	0.03	.98
Interaction of Time *	117.28	86.03	402.03	1.36	.174
Salience					
Weekend/-day	-1,633.85	378.85	402.04	-4.31	< .001*
Covariate					
Random effects					
Intercept	4,267,724	2,066			
Residual variance	12,645,677	3,556			
Number of Parameters	7				
-2LL	8,363.6				
AIC	8,377.6				
BIC	8,406.1				

* p-values are significant at α = .05.

Hypothesis 3: The Effects of Experience and Salience and Their Interaction on Steps per Day

Model Estimations. The model parameter estimates for the final model testing Hypothesis 3 can be found in Table 6. Overall, this model did not exhibit a significantly better model fit than the main models for Hypothesis 1 ($-2\Delta LL(3) = 5.98$, $p = .113$) or Hypothesis 2 ($-2\Delta LL(3) = 6.66$, $p = .084$). In the model testing Hypothesis 3, the interaction of tracking experience and time was significant, with an increasing difference of 251 steps taken less by the experienced trackers compared to novice trackers on each progressive day of the study period. Another significant model parameter was the covariate comparing the weekend to weekdays, with an average of approximately 1,632 less steps taken on the weekend compared to a weekday. There was no significant effect of the three-way interaction between time, tracking experience and salience in the main model. The simple slopes analysis revealed that there was no significant effect of time for the novice trackers, irrespective of whether they were in the low salience condition (see Table 6) or the high salience condition ($b = 6.49$, $SE = 75.76$, $t = 0.09$, $p = .93$). For experienced trackers in the low salience condition, there was a significant effect of time ($b = -277.29$, $SE = 108.86$, $t = -2.55$, $p = .01$), but not for those in the high salience condition ($b = 21.78$, $SE = 75.76$, $t = 0.29$, $p = .77$). Figure A 3 in Appendix A further illustrates the effects of the three-way interaction. These results contrast with Hypothesis 3, according to which experienced trackers in the low salience condition should exhibit the least pronounced fall in steps, while this is the only group for which a significant decline in steps was found. Thus, Hypothesis 3 was also rejected.

The model including all days contained both a significant effect of tracking experience, with experienced trackers taking roughly 2,024 steps more on the first day, and a significant interaction of experience and time, indicating that as one day of the study period passed, the decline in steps was 252 steps more in experienced trackers compared to novice trackers, but there was no significant change over time for novice trackers (see also Table A 7). The model without Days 1 and 15 also comprised a significant interaction of time and experience, with a decrease roughly 277 steps larger for experienced trackers than novice trackers each day, but no significant main effects (see also Table A 10). Although the models of interest for Hypothesis 3 included a significant interaction between time and tracking experience, it indicates an effect in the opposite direction of Hypothesis 1. Most importantly, neither of the models in the sensitivity analysis included a significant three-way interaction between time, tracking experience, and salience.

Table 6

Model Effects for Hypothesis 3 (Excluding Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	7,844.49	813.83	65.25	9.64	< .001*
Time	-25.95	75.74	427.83	-0.34	.732
Tracking experience	1,862.69	1,002	64.15	1.86	.068
Interaction of Time *	-251.34	126.13	379.70	-1.99	.047*
Tracking Experience					
Salience	-334.43	1,002.49	64.27	-0.33	.74
Interaction of Time *	32.44	103.63	429.65	0.31	.754
Salience					
Interaction of Time *	266.63	147.74	222.99	1.81	.073
Tracking Experience *					
Salience					
Weekend/-day	-1,631.54	376.47	401.28	-4.33	< .001*
Covariate					
Random effects					
Intercept	4,079,518	2,020			
Residual variance	12,475,844	3,532			
Number of Parameters	10				
-2LL	8,356.9				
AIC	8,376.9				
BIC	8,417.6				

* p-values are significant at $\alpha = .05$.

Effect Size Decomposition. Of the within-person outcome variance, 5.57% were attributable to $R_w^{2\,(f_1)}$, that is, the fixed effects of the Level-1 predictors study day and the weekend/-day covariate, the Level-1 by Level-2 cross-level interactions of study day and tracking experience as well as study day and salience, and the Level-1 by Level-2 part of the three-way interaction between study day, experience, and salience. The proportion of between-person outcome variance attributable to $R_b^{2\,(f_2)}$, that is, the fixed effects of the Level-2 predictors tracking experience and salience as well as the Level-2 by Level-2 part of the three-way interaction, was 14.93%. As indicated by $R_t^{2\,(f_1)}$, 4.09% of the total outcome variance may be accounted for by the fixed effects of the Level-1 predictors, the Level-1 by Level-2 cross-level interactions, and the Level-1 by Level-2 part of the three-way interaction. $R_t^{2\,(f_2)}$ suggests that 3.98% of the total outcome variance may be attributed to the fixed effects of the Level-2 predictors and the Level-2 by Level-2 part of the three-way interaction. The total outcome variance accounted for by the fixed effects of all predictors and interactions, indicated by $R_t^{2\,(f)}$, was 8.06%. A full breakdown of the set of R^2s is shown in Figure A 6 and in Table A 13 in Appendix A.

Effect Size Comparisons Between Models

Comparing the variance accounted for by the model of interest for Hypothesis 1 with an unconditional model of change (see also Table A 14 in Appendix A for model parameters and Table A 15 for the full set of R^2s) revealed that the proportion of total outcome variance attributable to the addition of the Level-2 predictor of prior physical activity tracking experience via $\Delta R_t^{2\,(f_2)}$ was 2.72% – 0% = 2.72% and the between-person outcome variance attributable to tracking experience via $\Delta R_b^{2\,(f_2)}$ was 10.91% – 0% = 10.91%. The interaction of tracking experience and time accounted for $\Delta R_t^{2\,(f_1)}$ = 3.41% – 3.33% = 0.08% of total outcome variance and $\Delta R_w^{2\,(f_1)}$ = 4.54% – 4.43% = 0.11% of within-person outcome variance. The model parameter of interest for Hypothesis 1 therefore accounted for only a rather small amount of variance.

When contrasting the variance accounted for by the model for Hypothesis 2 with the unconditional model of change, 0.86% of total outcome variance was attributable to the Level-2 predictor salience ($\Delta R_t^{2\,(f_2)}$ = 0.86 – 0%), as was 3.43% of the between-person outcome variance ($\Delta R_b^{2\,(f_2)}$ = 3.43% – 0%). The interaction of time and salience explained 0.3% of total outcome variance ($\Delta R_t^{2\,(f_1)}$ = 3.63% – 3.33 %) and 0.41% of within-person outcome variance ($\Delta R_w^{2\,(f_1)}$ = 4.84% – 4.43%). Therefore, the parameter of interest for the model for hypothesis 2 explained only little variance overall and within-person.

Comparing the model of interest for Hypothesis 3 with a reduced model lacking the three-way interaction between study day, experience, and salience (see also Table A 16 in Appendix A for model parameters and Table A 17 for the full set of R^2s) showed that 1.18% of the total outcome variance was attributable to this interaction, since it is comprised of both an interaction between Level-1 and Level-2 predictors, reflected in $\Delta R_t^{2\,(f_1)}$ = 4.09% – 3.79% =

0.3%, and an interaction between Level-2 predictors, reflected in $\Delta R_t^{2\,(f_2)}$ = 3.98% – 3.1% = 0.88%. Furthermore, of the within-person outcome variance, 0.52% were accounted for by the three-way interaction ($\Delta R_w^{2\,(f_1)}$ = 5.57% – 5.05%). Concerning between-person outcome variance, 2.5% were explained by the three-way interaction ($\Delta R_b^{2\,(f_2)}$ = 14.93% – 12.43%). This model parameter thus only appeared to play a relatively minor role in explaining variance.

Discussion

The present study aimed to investigate two potential mechanisms of measurement reactivity in physical activity assessed with objective physical activity monitors, namely the familiarity with physical activity tracking, which was presumed to differ between people with and without prior experience using personal fitness apps and trackers, and the salience of researcher observation. The following section is dedicated to a critical engagement with the current findings.

Interpretation of Results

Three hypotheses were meant to be investigated by the present study. First, it was suggested that measurement reactivity might be more pronounced in novice trackers, as evidenced by a steeper decline in steps over the study period. This was assumed because the act of measurement should carry more novelty for novice trackers. The results of the model specified to test this did not support the hypothesis. In the main model, for which the last day of the study period was excluded, there was no significant interaction between time and experience, which was the key parameter of interest, but also no significant effect of time, in neither experienced nor novice trackers, as the simple slope analysis revealed. In neither of the sensitivity analyses was there a significant interaction between time and tracking experience or a significant main effect of time. It thus appeared not only that there was no evidence for the proposed mechanism of measurement reactivity, but that there was also a lack of reactivity itself.

Second, it was hypothesised that higher salience of researcher observation would lead to stronger measurement reactivity, which should be reflected in a steeper decline in steps for participants in the high compared to the low salience condition. Again, the results did not support this hypothesis. The main model showed no significant effects of time, salience, or a time by salience interaction. This was further supported by the simple slope analysis, which showed no effect for time in either salience condition. In the first sensitivity analysis with all days, there was a significant effect of time, but no significant interaction with salience. In the second sensitivity analysis, there was a significant effect of time as well as a significant interaction between time and salience, but this interaction suggested that people in the low salience condition showed a steeper decline in steps over the study period, and not people in the high salience condition, as the hypothesis had proposed. Therefore, there was somewhat mixed evidence for the existence of measurement reactivity and differences in its effects between the salience groups in the present data, but what it indicated did not align with the propounded mechanism.

Third, the last hypothesis postulated a moderating effect of tracking experience on the interaction between time and salience, in that the effect of high salience on a steeper decline in steps should be more pronounced in novice trackers. In a way, this hypothesis combined the two previous ones and specified their interplay. Consequently, the results for the third hypothesis were also relevant to the first two hypotheses. The main model for the last hypothesis did not show a significant three-way interaction between time, salience, and experience. However, the simple slope analysis indicated that there was a significant effect of time for experienced trackers in the low salience condition, but not for any other group. Still, these results are not consistent with the third hypothesis, which proposed that the decline in steps over time should be the least pronounced in the only group for which a significant effect of time could, in fact, be attested. There was a significant effect of the interaction between time and tracking experience in the main model testing Hypothesis 3. This was also evident in both of the sensitivity analyses. This contrasts with the lack of a significant interaction between time and tracking experience in any of the models for the first hypothesis. It thus appears that this interaction might exist but only may become apparent when analysed in conjunction with salience. Similarly, a significant interaction of time by salience was only apparent in one of the sensitivity analyses for Hypothesis 2, but neither in the main model for hypothesis two nor the other sensitivity analysis. Still, the simple slope analysis for the model for Hypothesis 3 suggested that there may be such an interaction in experienced trackers. However, it is important to note the small sample sizes for these subgroup analyses. In particular, the group of experienced trackers in the low salience condition was comprised of only four people, so the robustness of the detected effect of time in this group is questionable. Moreover, the main model for Hypothesis 3 did not show a significantly better model fit than either of the other main models, so the greater complexity of the last model does not necessarily afford it additional explanatory power compared to the other two models.

Overall, there was no clear evidence of measurement reactivity across all groups of participants. Only when analysing simple slopes, a significant decline in steps could be attested in just one of four groups, comprising a rather low number of participants. Indeed, although the effects were not significant, there was even a positive trend for steps in both experienced and novice trackers in the high salience condition evident in the simple slopes analysis. The basic expectation underlying all hypotheses that the number of steps taken would fall over the study period for all participants, with only differences in the extent of the decline, was not fulfilled. Since the anticipated pattern of measurement reactivity could not be clearly and consistently attested, it is also difficult to draw conclusion about the potential mechanisms of measurement reactivity, as was the intent of the present study. Indeed, the one relatively confident inference that can be drawn from the current data was that physical activity in terms of steps taken per day differed between weekdays (Monday to Friday) and weekends (Saturday and Sunday), with less steps taken on the weekends, as the only model parameter consistently significant across all nine models was the covariate coding for this comparison.

In this context, the operationalisation of measurement reactivity for this study calls for reflection. Most studies investigating measurement reactivity in physical activity either search for a significant difference between Day 1 or Day 1 and 2 and the rest of the days of the study period or a significant change over time (König et al., 2022). The latter approach was taken in the current study, defining measurement reactivity as a linear change during the study period. As previously mentioned, there were no obvious points of inflection or polynomial trends apparent in a visual inspection of the current data, which would indicate the inappropriateness of applying a linear trend to the data. However, according to the descriptive statistics detailing the progression of steps over the study period for the whole sample and within each tracking experience and salience group, there did seem to be a tendency for the number of steps to slightly meander. Compared to other research projects, the current study employed a relatively long study protocol, requiring participants to wear the accelerometer for 14 days. Most frequently in observational studies, physical activity is measured for 7 days (Evenson et al., 2022), as was also the case in 10 out of 25 (40%) of the studies on measurement reactivity in physical activity included by König et al. (2022). Assuming measurement reactivity is present chiefly in the first few days of measurement, it is possible that in conjunction with a linear trend, the utilisation of a slightly shorter study period is more suitable for the detection of reactivity. It could, moreover, be interesting to directly compare the results of analyses employing a linear trend and a contrast between specific days of the study period, like Arigo

& König (2024). Although such an explicit comparison was not conducted here, it nevertheless appears somewhat unlikely in light of the descriptively slightly zigzagging pattern of steps that Day 1 or Day 1 and 2 taken together would indeed differ significantly from all other days.

The finding that no robust measurement reactivity effects were present in the current analyses is quite surprising. There are several aspects of the present study which would suggest measurement reactivity to be particularly likely to arise. These include the fact that participants were generally aware that the outcome of interest was physical activity, which is a widely recognised health behaviour, with more physical activity conventionally considered more socially desirable (French et al., 2021; Kulnik et al., 2025). Moreover, a sample predominantly consisting of students was recruited, and somewhat restrictive inclusion criteria were applied, with participants being excluded if they did not fit into either of the predefined categories of experienced or novice trackers; both of which also would point to a higher probability of measurement reactivity (French et al., 2021). Of course, the study features likely to elicit measurement reactivity compiled by French et al. (2021) were chiefly based on research employing self-report measures for a range of different behaviours and thus only constitute indirect evidence in the context of physical activity measured with objective sensors. Still, as French et al. argued, there do not seem to be convincing reasons to assume that these features should not have a similar impact in studies beyond those utilising self-report measures. Furthermore, another choice for the present study which should have made the detection of measurement reactivity more likely was that steps specifically were chosen as an indicator of physical activity, as other indicators like moderate-to-vigorous physical activity may be less prone to reactivity (König et al., 2022). Theoretically, these factors should have been a strength in the design of the current study.

Nevertheless, not all studies find evidence of measurement reactivity in physical activity measured with objective motion sensors. Even in the systematic review and meta-analysis by König et al. (2022), which broadly showed that measurement reactivity in this context can exist, a sizeable minority of five out of 15 studies (33%) on physical activity in terms of steps did not find significant findings. Moreover, of these 15 studies, only four experimental studies could be included in the meta-analysis, which was thus based on a relatively limited amount of evidence, while still exhibiting considerable heterogeneity. A different meta-analysis on changes in objectively measured physical activity within control groups in primary care physical activity intervention studies also found no significant changes in steps per day (Freene et al., 2020). However, the extent to which the results of this meta-analysis are comparable to the present study is unclear, as an increase in physical activity from a baseline measurement to follow-up was expected by the authors, after a median intervening duration of 24 weeks. Additionally, the sample of Freene et al. (2020) was different from the present one in several aspects, including the age, BMI, and the presence or risk of chronic disease in many participants. Still, measurement reactivity might not always occur in such a prominent way as to be detectable and thus create noticeable bias.

Even so, it seems premature to dismiss the possibility of measurement reactivity in the current research project. Considering that the existing evidence indicates that measurement reactivity on average tends to produce medium effects at most (König et al., 2022), it is conceivable that the present analyses were simply based on too small a sample, with only 31 participants, to achieve the necessary statistical power to detect an existing but relatively small effect. In this context, the large variability in the number of steps taken between the participants in the current analyses should be noted, as evident in the substantial standard deviations on most days of the study period both in the whole sample and each of the tracking experience and salience conditions. Moreover, the intraclass correlation was relatively small, indicating that a larger proportion of the variance was attributable to within-person differences than between-person differences. This considerable variance might have led to a muddying of effects. Additionally, a further factor could have interfered with the detection of a linear trend in steps over time. A sizeable portion of the participants, namely 11 people (35%), reported stressful life events during the study period at the second session in the laboratory. Unfortunately, no further information was recorded on what exactly those events where, if and how they impacted the participant's life, and when precisely they occurred. Yet it is possible if these stressful life events took place at some time during the middle of the study period, leading to a drop in physical activity, from which participants later returned to their usual activity levels, this could spoil a clear linear trend of steps decreasing more or less continually from the beginning to the end of the study period.

Potential Mechanisms of Measurement Reactivity

No clear evidence for the presence of measurement reactivity could be found in the present study, preventing definitive conclusions about its potential mechanisms. Still, the attempt to fill the existing knowledge gap concerning mechanisms of measurement reactivity should be considered a strength of the present study. Some reflections on this novel paradigm could help guide future research. Concerning the effect of the salience of researcher observation, it is somewhat tricky to make sense of the present findings. There were mixed results on whether a significant interaction between time and salience existed, but what evidence there was seemed to indicate a stronger measurement reactivity effect in terms of a significant decline in steps in the participants in the low salience condition. While it is theoretically conceivable that the measurement reactivity in the high salience group could be so pronounced as to be present throughout the whole study period, with a decline in steps only occurring after more than 14 days, this seems extremely unlikely in light of the existing literature on the duration of measurement reactivity (Arigo & König, 2024; Clemes & Deans, 2012; French et al., 2021; König et al., 2022). A more likely interpretation is that these differences are better explained by factors other than the salience manipulation, outside of the study's purview.

Indeed, although the induction of differential salience utilised in the present study was intended to be more powerful than that of König et al. (2025), by delivering it in-person to research participants, it is still possible that two additional sentences explaining the study procedure in more detail did not effectively increase the salience of being observed for participants in the high salience condition. It would be interesting for future research to create even more marked salience manipulations to further investigate this potential mechanism, for example by utilising differential reminder protocols to prompt participants' awareness of their study participation. However, such a manipulation might lead to diversity in data quality if, for instance, those participants with more frequent reminders wear their motions sensors for longer periods of time, which would need to be taken into consideration.

In terms of tracking experience as a proxy for familiarity with, or decreased novelty of, physical activity monitoring, there are several points to discuss. First, the assignment of participants to the experienced and novice tracking groups could potentially be seen as somewhat problematic. People had to have used a fitness app or tracker regularly for at least 14 days prior to enrolment in order to be classified as an experienced tracker. This cut-off was somewhat arbitrarily based on the pragmatic need for experienced trackers to be able to report their step counts for the 2 weeks before the study period. However, since the average duration of use was about 34 months, and experienced trackers self-reported a high consistency of tracker use, it can reasonably be assumed that the people in the experienced tracking group did, in fact, have ample experience with their personal trackers. For people to be categorised as novice trackers, they could not ever have used any fitness app or tracker. It is possible that some people either misunderstood the designation "fitness app or tracker" to mean not hardware and software to monitor physical activity, but, for example, apps that provide guided workouts or running programmes, or, alternatively, that they forgot that they had used a physical activity monitor before. Although the eligibility criteria were specified during the recruitment process and participant were moreover personally questioned by an experimenter on their tracking experience status at the beginning of the first session, there is nonetheless no guarantee of accurate group assignment since it was based on participants' self-reports, with all of the potential caveats that come with this type of information. Any person who did not fit into the predefined categories of experienced and novice physical activity trackers, for instance because they did not currently use a fitness tracker or app but had done so in the past or only occasionally used such a device was excluded from participation. This procedure was meant to yield more easily interpretable results because of an anticipated stronger contrast between the two tracking experience groups. From an alternative position, this could have led to the discarding of useful information on the effects of varying degrees of tracking experience. Future research will have to consider how to weigh the significance of these arguments.

Second, although the experienced trackers had been using personal fitness apps and trackers for several years on average, is unclear to what extent the familiarity with personal devices actually carried over to the devices used in the study, and how far the novelty associated with physical activity monitoring was indeed reduced for experienced compared to novice trackers. After all, the actual accelerometer used in the study was similarly unfamiliar to both experienced and novice trackers. It could be valuable for further research utilising a similar paradigm to attempt to quantify the degree of the participants' perceived novelty of the specific study's measurement protocol.

Third, other variables might be associated with tracking experience, not all of which were accounted for in the current study. Sociodemographic variables like gender, age, education, as well as overweight have all been associated with physical activity, with men (compared to women), younger people, those with higher levels of education, and those with normal weight tending to be more active (Bauman et al., 2012). Gender, age, highest level of education, and BMI were assessed and, except for experienced trackers being older on average, no differences were found between participants in either the tracking experience or salience groups, which is not surprising considering the overall quite homogenous sample. Gender, age, and education level have moreover been linked to the use of health apps and activity monitors more broadly (Carroll et al., 2017; Cornejo Müller et al., 2020; Omura et al., 2017). Interestingly, while the associations concerning the use of health apps and activity monitors with age and education are consistent with those on physical activity, women have actually been described as more likely to use health apps and activity monitors (Carroll et al., 2017; Omura et al., 2017), though there may be gender differences in the exact kinds of applications used (Cornejo Müller et al., 2020). Users of health apps and activity monitors are also more likely to report meeting the US national physical activity guidelines of a minimum of 150 minutes of moderate physical activity per week (Carroll et al., 2017; Omura et al., 2017). Although not present in the main analyses, there were significant main effects of experience in the sensitivity analyses with all days in the models for the first and third hypotheses, suggesting a somewhat inconsistent tendency for experienced trackers in the present sample to perform more physical activity. Users of health apps and physical activity monitors have also been found to be more likely to hold personal health or fitness goals like wanting to increase physical activity or lose weight (Carroll et al., 2017; Jin et al., 2022). However, it is not clear whether people with preexisting higher levels of physical activity are especially drawn to health apps and monitors or whether the use of such apps and monitors may impact the user's physical activity over time (Carroll et al., 2017; Jin et al., 2022), or there may be even more complicated causal mechanisms at play involving other variables.

Adding to this tangle of interrelated variables, health literacy and in particular eHealth literacy have been linked to the use of health apps and fitness monitors (Cornejo Müller et al., 2020). Similarly, health consciousness has been found to have a direct effect on the extent of health app use, with health app use efficacy mediating the effect of health information orientation and eHealth literacy on extent of health app use (Cho et al., 2014). Personal fitness goals, health literacy and health consciousness were not part of the current study but may play a role in physical-activity related differences between people with and without experience with physical activity tracking. This could relate to both general differences in physical activity levels as well as its trajectory over time. As speculated by French et al. (2021), people with personal ulterior motivations to engage with a study and/or greater health awareness may be particularly prone to measurement reactivity. In the present analyses, the evidence for differential measurement reactivity between experienced and novice trackers was somewhat inconsistent, but tendentially, experienced trackers appeared to exhibit more measurement reactivity. Future research on measurement reactivity in the context of physical activity may benefit from further investigating the connections between a variety of variables linked to tracking experience, including those mentioned here.

Accelerometer-Derived Steps as Physical Activity Estimates

Moving on from the discussion of the present findings directly related to measurement reactivity and it potential mechanisms, there are also more methodological considerations to take into account when interpreting this study's results. One of these relates to the way physical activity was operationalised. Although objective physical activity measures like accelerometers are often deemed to be more accurate than subjective measures (Ainsworth et al., 2015; Dowd et al., 2018), they nevertheless also only represent estimates of the variable of interest (Burchartz et al., 2020). Physical activity is a multi-faceted behaviour (Ainsworth et al., 2015; Strath et al., 2013), and no one measure is capable of capturing all potentially relevant aspects (Ainsworth et al., 2015; Reichert et al., 2020; Sattler et al., 2021; Skender et al., 2016). In particular, steps per day were chosen as the key physical activity indicator for this study. Steps represent a basic metric of ambulatory activity (Bassett et al., 2017), which chiefly comprises the kinds of physical activity accelerometers are best suited to capture (Ainsworth et al., 2015; Pedišić & Bauman, 2015). Differently put, there are a variety of types of physical activity which may be poorly reflected by accelerometer-derived step count estimates. Consequently, this delimits the scope of the present research.

Additionally, steps were determined by the movisens StepCount algorithm and wear time was determined based on the NonWearTime algorithm. The use of proprietary algorithms for accelerometer data processing has been criticised because it hampers the replicability of findings and the comparability between studies using different algorithms (Ainsworth et al., 2015; Burchartz et al., 2020; Evenson et al., 2022; Migueles et al., 2017). Compounding this issue is the fact that the accelerometer brand utilised for this project, movisens, is somewhat younger than other accelerometer brands commonly employed. The company was founded in 2009 (movisens, n.d.), while the first observational study using accelerometry was published 10 years prior, according to a scoping review, which also saw a pronounced rise in research in the late 2000s (Evenson et al., 2022). This review also detailed 25 different accelerometer brands, but movisens was not on the list, suggesting that accelerometers by this brand have not been used as widely as others. Although one study deemed physical activity estimates by movisens Move 4 highly comparable to those of other accelerometers more widely used in physical activity research when data processing was done in an identical manner (Migueles et al., 2022), it not certain to which extent this finding is applicable more generally, including to situations where proprietary algorithms have been employed.

It is thus valuable to evaluate how the estimates of steps per day derived from the present data contrast to other research. Over Days 1 to 14 of the study period, the descriptive mean number of steps taken by the participants in this study was 7,735 (*SD* = 4,214). This would correspond to the "somewhat active" group (7,500 – 9,999 steps) of the classification by Tudor-Locke and Bassett (2004). The present step estimate is lower than the mean of 9,124 steps per day among adults aged 40 years and over in the 2003-2006 NHANES, obtained with the ActiGraph 7164 accelerometer, as reported by an analysis of steps per day and all-cause mortality (Saint-Maurice et al., 2020). However, an earlier analysis of 2005-2006 NHANES data found an estimated mean of 9,676 steps and deemed it unrealistically high, and thus censored out steps taken at an intensity of less than 500 activity counts per minute, yielding mean of 6,540 steps per day judged to be more likely to be accurate (Tudor-Locke et al., 2009), which is lower than the present sample mean. A meta-analysis pooling data from 47,471 adults from both published and unpublished cohort studies on steps per day and their association with all-cause mortality using several different brands of accelerometers, though not movisens Move 4, found a median of 7,803 steps (interquartile range 5,377 – 10,352) in adults under 60 years of age over all included studies (Paluch et al., 2022), which closely mirrors the results of the current analysis. Of course, the studies mentioned here employed samples differing in a variety of variables, including, but not limited to, age, geographical location, and year(s) that the study was conducted in. Still, in spite of the range of step count estimates variously reported, it appears the current step estimates may be considered broadly comparable to those derived from other accelerometer brands and models.

Exclusion of Participants

A further point of discussion is the choice of wear time duration required of participants to be included in the present analyses. It has been recommended that for 1 week of physical activity monitoring, a minimum of 4 days with at least 10 hours of wear time should be accrued by participants (Migueles et al., 2017). Indeed, the most commonly implemented wear time protocol in observational studies on physical activity is 1 week of observation with at least 4 days of at least 10 hours of wear time (Evenson et al., 2022). Since the study period for the current research project was 14 days, the number of required days was set to 8 days for the present analyses. As previously described, a close connection has been found to exist between the minimum wear time criterium, the estimated amount of physical activity, and ensuing the sample size (Herrmann et al., 2014; Keadle et al., 2014). This applied also to the present study, where a less severe wear time requirement could have led to a larger sample, instead of having to exclude roughly a third of potentially eligible participants. Moreover, there may be systematic differences between participants who meet the minimum wear time criterium and those who do not (Loprinzi et al., 2013), suggesting a strong dependency of derived physical activity estimates on the precise sample under investigation. However, lowering the minimum wear time either in terms of valid hours per day or valid days during the study period would have been associated with some degree of data quality reduction. Although the first few days of the study period were the most important since measurement reactivity is most likely to manifest at this point (König et al., 2022), the other days are also relevant to examine whether changes in physical activity also occur at other times, and to compare the magnitude of these changes. Moreover, in terms of the required hours per day, considering that adults report spending an average of 7.8 hours (*SD* = 0.9) a day in bed (Kocevska et al., 2021), exacting 10 hours of wear time during waking hours of participants actually seems a relatively low hurdle. Indeed, a low number of hours of wear time could be deemed particularly problematic since the derived estimates are unlikely to accurately reflect the true values of a given person's physical activity that day, whereas a day with no wear time simply adds no information. This is especially true for analyses like the present, which relied on multilevel modelling, for which missing data is no issue. So while there may be some degree of arbitrariness associated with the present decision, it nevertheless could be considered a pragmatic approach trying to balance both concerns for data quality and sample size.

Exclusion of Days

Another pivotal decision in the present study was the exclusion of the last day of the study period for the main models to test the hypotheses. This was mainly based on the observation that participants displayed very little wear time on that day and the associated step estimates were also much lower than on any other day. To avoid an artificially significant effect of time caused mainly by the plunge of steps on the last day, rather than significant changes at the beginning of the study period, the last day was only included as part of a sensitivity analysis. The necessity for taking this approach was partly based in the study procedure, where participants were handed the accelerometers in person during the day and the recording was started immediately, resulting in the last day constituting only part of a day. Of course, a similar procedural issue was also present for the first day of the study period, which did not reflect a full day either. For this study, the first day was moreover excluded as part of a sensitivity analysis. Consequently, some effects were shown to be more robust than others. Conducting these sensitivity analyses and comparing the effects over three different definitions of the period of interest may be considered a strength of the present study.

Even so, alternative procedures would also have been possible. First, the estimates for steps could have been adjusted for wear time. However, this would have involved a number of further decisions with some degree of subjectivity. For example, it would have included making assumptions on what precisely the duration of wear time for a realistic full day would be, that is, determining the weights with which to adjust the raw values. This is partly based on the fact that there was considerable variation in the descriptive wear time accrued by participants both within-person and between-person, even apart from the first and last days of the study period. Second, it could be potentially questionable to generalise physical activity estimates beyond the recorded wear time because there may be systematic differences in physical activity between times when the accelerometer is worn and when it is not. It could be that a sensor is deemed impractical or uncomfortable (Jin et al., 2022), making it more likely to be taken off for kinds of physical activity where it might be considered a burden. Alternatively, it appears a reasonable conjecture that participants may be more likely not to wear a motion sensor when they do not perform any physical activity because the obligation to wear the sensor may be less salient in this case or because participants suppose it to be less important to record non-activity. Third, it would need to be considered which days should be adjusted, i.e. only the last day, or also the first day, or potentially other days as well during which some kind of minimum wear time was not met. An alternative to the adjustment of physical activity estimates based on wear time would be to define a day as each consecutive 24-hour period after the start of recording rather than based on dates as in the present study. A further option would be to start the measurement at a later time, for example at midnight, and not immediately when the participant receives the motion sensor. Then again, participants may exhibit increased physical activity due to measurement reactivity already during the period where they have received a physical activity monitor but it has not yet commenced recording. All of these possibilities could be further explored and contrasted by future research.

Conclusion

This study intended to examine two potential mechanisms of measurement reactivity in objectively measured physical activity, namely the familiarity with physical activity tracking in terms of experience with personal fitness apps and trackers, and the salience of researcher observation. Since no consistent effects of measurement reactivity were found throughout the sample, no definitive conclusions can be drawn on its underlying mechanisms. The present analyses were based on the available data at the time of writing, but the larger research project is still ongoing at this time. It will be interesting to see whether the current findings can be replicated with an extended sample. More generally, future research on measurement reactivity and its potential mechanisms will need to employ samples large enough to determine under which circumstances and for which people measurement reactivity may or may not arise. For a successful evaluation of the processes underlying measurement reactivity, inducing noticeable measurement reactivity in all intended groups of the sample is a necessary prerequisite. What the present piece of research should indicate is not that measurement reactivity is, in fact, not necessarily likely to cause bias, but instead that the specific conditions under which it arises are still relatively poorly understood as of yet.

References

Ainsworth, B., Cahalin, L., Buman, M., & Ross, R. (2015). The current state of physical activity assessment tools. *Progress in Cardiovascular Diseases*, *57*(4), 387–395. https://doi.org/10.1016/j.pcad.2014.10.005

Ammar, A., Trabelsi, K., Hermassi, S., Kolahi, A.-A., Mansournia, M. A., Jahrami, H., Boukhris, O., Boujelbane, M. A., Glenn, J. M., Clark, C. C. T., Nejadghaderi, A., Puce, L., Safiri, S., Chtourou, H., Schöllhorn, W. I., Zmijewski, P., & Bragazzi, N. L. (2023). Global disease burden attributed to low physical activity in 204 countries and territories from 1990 to 2019: Insights from the Global Burden of Disease 2019 Study. *Biology of Sport*, *40*(3), 835–855. https://doi.org/10.5114/biolsport.2023.121322

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

Bassett, D. R., Toth, L. P., LaMunion, S. R., & Crouter, S. E. (2017). Step counting: A review of measurement considerations and health-related applications. *Sports Medicine*, *47*(7), 1303–1315. https://doi.org/10.1007/s40279-016-0663-1

Batek, I. (2025). *Tracking the Trackers: Messreaktivität bei erfahrenen Fitnesstracker-Nutzer*innen* [Unpublished master's thesis]. University of Vienna.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., Krivitsky, P. N., Tanaka, E., Jagan, M., & Boylan, R. D. (2025). *lme4: Linear Mixed-Effects Models using „Eigen" and S4* (Version 1.1-37) [Software]. https://CRAN.R-project.org/package=lme4

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

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

Bendau, A., Petzold, M. B., Kaminski, J., Plag, J., & Ströhle, A. (2024). Exercise as treatment for “stress-related” mental disorders. *Current Neuropharmacology*, *22*(3), 420–436. https://doi.org/10.2174/1570159X22666230927103308

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

Burchartz, A., Anedda, B., Auerswald, T., Giurgiu, M., Hill, H., Ketelhut, S., Kolb, S., Mall, C., Manz, K., Nigg, C. R., Reichert, M., Sprengeler, O., Wunsch, K., & Matthews, C. E. (2020). Assessing physical behavior through accelerometry – State of the science, best practices and future directions. *Psychology of Sport and Exercise*, *49*, Article 101703. https://doi.org/10.1016/j.psychsport.2020.101703

Carroll, J. K., Moorhead, A., Bond, R., LeBlanc, W. G., Petrella, R. J., & Fiscella, K. (2017). Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *Journal of Medical Internet Research*, *19*(4), Article e125. https://doi.org/10.2196/jmir.5604

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports*, *100*(2), 126–131.

Chen, K. Y., & Bassett, D. R. J. (2005). The technology of accelerometry-based activity monitors: Current and future. *Medicine & Science in Sports & Exercise*, *37*(11), S490–S500. https://doi.org/10.1249/01.mss.0000185571.49104.82

Chiesa, M., & Hobbs, S. (2008). Making sense of social research: How useful is the Hawthorne Effect? *European Journal of Social Psychology*, *38*(1), 67–74. https://doi.org/10.1002/ejsp.401

Cho, J., Park, D., & Lee, H. E. (2014). Cognitive factors of using health apps: Systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. *Journal of Medical Internet Research*, *16*(5), Article e125. https://doi.org/10.2196/jmir.3283

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., Matchett, N., & Wane, S. L. (2008). 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

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

Cornejo Müller, A., Wachtler, B., & Lampert, T. (2020). Digital Divide – Soziale Unterschiede in der Nutzung digitaler Gesundheitsangebote [Digital Divide—Social Differences in the use of digital health services]. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *63*(2), 185–191. https://doi.org/10.1007/s00103-019-03081-y

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

Ding, D., Lawson, K. D., Kolbe-Alexander, T. L., Finkelstein, E. A., Katzmarzyk, P. T., Mechelen, W. van, & Pratt, M. (2016). The economic burden of physical inactivity: A global analysis of major non-communicable diseases. *The Lancet*, *388*(10051), 1311–1324. https://doi.org/10.1016/S0140-6736(16)30383-X

Dössegger, A., Ruch, N., Jimmy, G., Braun-Fahrländer, C., Mäder, U., Hänggi, J., Hofmann, H., Puder, J. J., Kriemler, S., & Bringolf-Isler, B. (2014). Reactivity to accelerometer measurement of children and adolescents. *Medicine & Science in Sports & Exercise*, *46*(6), 1140. https://doi.org/10.1249/MSS.0000000000000215

Dowd, K. P., Szeklicki, R., Minetto, M. A., Murphy, M. H., Polito, A., Ghigo, E., van der Ploeg, H., Ekelund, U., Maciaszek, J., Stemplewski, R., Tomczak, M., & Donnelly, A. E. (2018). A systematic literature review of reviews on techniques for physical activity measurement in adults: A DEDIPAC study. *International Journal of Behavioral Nutrition and Physical Activity*, *15*(1), Article 15. https://doi.org/10.1186/s12966-017-0636-2

Ekelund, U., Tarp, J., Steene-Johannessen, J., Hansen, B. H., Jefferis, B., Fagerland, M. W., Whincup, P., Diaz, K. M., Hooker, S. P., Chernofsky, A., Larson, M. G., Spartano, N., Vasan, R. S., Dohrn, I.-M., Hagströmer, M., Edwardson, C., Yates, T., Shiroma, E., Anderssen, S. A., & Lee, I.-M. (2019). Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: Systematic review and harmonised meta-analysis. *BMJ*, *366*, Article l4570. https://doi.org/10.1136/bmj.l4570

Evenson, K. R., Scherer, E., Peter, K. M., Cuthbertson, C. C., & Eckman, S. (2022). Historical development of accelerometry measures and methods for physical activity and sedentary behavior research worldwide: A scoping review of observational studies of adults. *PLOS ONE*, *17*(11), Article e0276890. https://doi.org/10.1371/journal.pone.0276890

Fiedler, J., Eckert, T., Burchartz, A., Woll, A., & Wunsch, K. (2021). Comparison of self-reported and device-based measured physical activity using measures of stability, reliability, and validity in adults and children. *Sensors*, *21*(8), Article 8. https://doi.org/10.3390/s21082672

Firth, J., Solmi, M., Wootton, R. E., Vancampfort, D., Schuch, F. B., Hoare, E., Gilbody, S., Torous, J., Teasdale, S. B., Jackson, S. E., Smith, L., Eaton, M., Jacka, F. N., Veronese, N., Marx, W., Ashdown-Franks, G., Siskind, D., Sarris, J., Rosenbaum, S., … Stubbs, B. (2020). A meta-review of "lifestyle psychiatry": The role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World Psychiatry*, *19*(3), 360–380. https://doi.org/10.1002/wps.20773

Freene, N., Davey, R., Sathiyakumar, R., & McPhail, S. M. (2020). Can physical activity measurement alone improve objectively-measured physical activity in primary care? A systematic review and meta-analysis. *Preventive Medicine Reports*, *20*, Article 101230. https://doi.org/10.1016/j.pmedr.2020.101230

French, D. P., Miles, L. M., Elbourne, D., Farmer, A., Gulliford, M., Locock, L., Sutton, S., & McCambridge, J. (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., & 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

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (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

Herrmann, S. D., Barreira, T. V., Kang, M., & Ainsworth, B. E. (2014). Impact of accelerometer wear time on physical activity data: A NHANES semisimulation data approach. *British Journal of Sports Medicine*, *48*(3), 278–282. https://doi.org/10.1136/bjsports-2012-091410

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.

Jakicic, J. M., Kraus, W. E., Powell, K. E., Campbell, W. W., Janz, K. F., Troiano, R. P., Sprow, K., Torres, A., & Piercy, K. L. (2019). Association between bout duration of physical activity and health: Systematic review. *Medicine & Science in Sports & Exercise*, *51*(6), 1213–1219. https://doi.org/10.1249/MSS.0000000000001933

Jin, D., Halvari, H., Maehle, N., & Olafsen, A. H. (2022). Self-tracking behaviour in physical activity: A systematic review of drivers and outcomes of fitness tracking. *Behaviour & Information Technology*, *41*(2), 242–261. https://doi.org/10.1080/0144929X.2020.1801840

Keadle, S. K., Shiroma, E. J., Freedson, P. S., & Lee, I.-M. (2014). Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study. *BMC Public Health*, *14*(1), Article 1210. https://doi.org/10.1186/1471-2458-14-1210

Klenk, J., Peter, R. S., Rapp, K., Dallmeier, D., Rothenbacher, D., Denkinger, M., Büchele, G., Becker, T., Böhm, B., Scharffetter-Kochanek, K., Stingl, J., Koenig, W., Riepe, M., Peter, R., Geiger, H., Ludolph, A., von Arnim, C., Nagel, G., Weinmayr, 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), Article 18. https://doi.org/10.1186/s11556-019-0226-1

Kocevska, D., Lysen, T. S., Dotinga, A., Koopman-Verhoeff, M. E., Luijk, M. P. C. M., Antypa, N., Biermasz, N. R., Blokstra, A., Brug, J., Burk, W. J., Comijs, H. C., Corpeleijn, E., Dashti, H. S., de Bruin, E. J., de Graaf, R., Derks, I. P. M., Dewald-Kaufmann, J. F., Elders, P. J. M., Gemke, R. J. B. J., … Tiemeier, H. (2021). Sleep characteristics across the lifespan in 1.1 million people from the Netherlands, United Kingdom and United States: A systematic review and meta-analysis. *Nature Human Behaviour*, *5*(1), 113–122. https://doi.org/10.1038/s41562-020-00965-x

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

König, L. M., Koch, T. J. S., & Arigo, D. (2024). *Why does reactivity to physical activity measurement occur? Empirical test of potential mechanisms.* Retrieved from https://osf.io/s9vnh/?view_only=e2aab1583c89418da7ec35f19b6e24e1

König, L. M., Pasko, K., Baga, K., Harsora, R., & Arigo, D. (2025). Isolating the role of researcher observation on reactivity to the measurement of physical activity. *Applied Psychology: Health and Well-Being*, *17*(1), Article e12630. https://doi.org/10.1111/aphw.12630

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), Article e55. https://doi.org/10.2196/mhealth.8261

Kulnik, S. T., Leysen, D., & Pfannerstill, F. (2025). Unwanted bias or promising intervention effect? Dealing with measurement reactivity in controlled trials of physical activity interventions. *European Journal of Cardiovascular Nursing*, Corrected Proof. https://doi.org/10.1093/eurjcn/zvaf121

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2022). *lmerTest: Tests in Linear Mixed Effects Models* (Version 3.1-3) [Software]. https://cran.r-project.org/web/packages/lmerTest/index.html

Long, J. A. (2025). *Interactions: Comprehensive, user-friendly toolkit for probing interactions* (Version 1.2.0) [Software]. https://CRAN.R-project.org/package=interactions

Loprinzi, P. D., Cardinal, B. J., Crespo, C. J., Brodowicz, G. R., Andersen, R. E., & Smit, E. (2013). Differences in demographic, behavioral, and biological variables between those with valid and invalid accelerometry data: Implications for generalizability. *Journal of Physical Activity and Health*, *10*(1), 79–84.

Matthews, C. E. (2005). Calibration of accelerometer output for adults. *Medicine & Science in Sports & Exercise*, *37*(11), S512–S522. https://doi.org/10.1249/01.mss.0000185659.11982.3d

McCambridge, J., Kypri, K., & Elbourne, D. (2014). Research participation effects: A skeleton in the methodological cupboard. *Journal of Clinical Epidemiology*, *67*(8), 845–849. https://doi.org/10.1016/j.jclinepi.2014.03.002

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

McVeigh, J. A., Winkler, E. A. H., Howie, E. K., Tremblay, M. S., Smith, A., Abbott, R. A., Eastwood, P. R., Healy, G. N., & Straker, L. M. (2016). Objectively measured patterns of sedentary time and physical activity in young adults of the Raine study cohort. *International Journal of Behavioral Nutrition and Physical Activity*, *13*, Article 41. https://doi.org/10.1186/s12966-016-0363-0

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

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., Labayen, I., Ruiz, J. R., & Ortega, F. B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine*, *47*(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0

Migueles, J. H., Molina-Garcia, P., Torres-Lopez, L. V., Cadenas-Sanchez, C., Rowlands, A. V., Ebner-Priemer, U. W., Koch, E. D., Reif, A., & Ortega, F. B. (2022). Equivalency of four research-grade movement sensors to assess movement behaviors and its implications for population surveillance. *Scientific Reports*, *12*(1), Article 5525. https://doi.org/10.1038/s41598-022-09469-2

Montoye, A. H. K., Moore, R. W., Bowles, H. R., Korycinski, R., & Pfeiffer, K. A. (2018). Reporting accelerometer methods in physical activity intervention studies: A systematic review and recommendations for authors. *British Journal of Sports Medicine*, *52*(23), 1507–1516. https://doi.org/10.1136/bjsports-2015-095947

Mora-Gonzalez, J., Gould, Z. R., Moore, C. C., Aguiar, E. J., Ducharme, S. W., Schuna, J. M., Barreira, T. V., Staudenmayer, J., McAvoy, C. R., Boikova, M., Miller, T. A., & Tudor-Locke, C. (2022). A catalog of validity indices for step counting wearable technologies during treadmill walking: The CADENCE-adults study. *International Journal of Behavioral Nutrition and Physical Activity*, *19*(1), Article 117. https://doi.org/10.1186/s12966-022-01350-9

movisens. (n.d.). *Unternehmen*. https://www.movisens.com/de/unternehmen/

Omura, J. D., Carlson, S. A., Paul, P., Watson, K. B., & Fulton, J. E. (2017). National physical activity surveillance: Users of wearable activity monitors as a potential data source. *Preventive Medicine Reports*, *5*, 124–126. https://doi.org/10.1016/j.pmedr.2016.10.014

Paluch, A. E., Bajpai, S., Bassett, D. R., Carnethon, M. R., Ekelund, U., Evenson, K. R., Galuska, D. A., Jefferis, B. J., Kraus, W. E., Lee, I.-M., Matthews, C. E., Omura, J. D., Patel, A. V., Pieper, C. F., Rees-Punia, E., Dallmeier, D., Klenk, J., Whincup, P. H., Dooley, E. E., … Fulton, J. E. (2022). Daily steps and all-cause mortality: A meta-analysis of 15 international cohorts. *The Lancet Public Health*, *7*(3), e219–e228. https://doi.org/10.1016/S2468-2667(21)00302-9

Pedišić, Ž., & Bauman, A. (2015). Accelerometer-based measures in physical activity surveillance: Current practices and issues. *British Journal of Sports Medicine*, *49*(4), 219–223. https://doi.org/10.1136/bjsports-2013-093407

Reichert, M., Giurgiu, M., Koch, E. D., 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*, *50*, Article 101742. https://doi.org/10.1016/j.psychsport.2020.101742

Reiner, M., Niermann, C., Jekauc, D., & Woll, A. (2013). Long-term health benefits of physical activity – A systematic review of longitudinal studies. *BMC Public Health*, *13*(1), Article 813. https://doi.org/10.1186/1471-2458-13-813

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184

Rights, J. D., & Sterba, S. K. (2021). Effect size measures for longitudinal growth analyses: Extending a framework of multilevel model R-squareds to accommodate heteroscedasticity, autocorrelation, nonlinearity, and alternative centering strategies. *New Directions for Child and Adolescent Development*, 65–110. https://doi.org/10.1002/cad.20387

Saint-Maurice, P. F., Troiano, R. P., Bassett, D. R., Jr, Graubard, B. I., Carlson, S. A., Shiroma, E. J., Fulton, J. E., & Matthews, C. E. (2020). Association of daily step count and step intensity with mortality among US adults. *JAMA*, *323*(12), 1151–1160. https://doi.org/10.1001/jama.2020.1382

Sattler, M. C., Ainsworth, B. E., Andersen, L. B., Foster, C., Hagströmer, M., Jaunig, J., Kelly, P., Iii, H. W. K., Matthews, C. E., Oja, P., Prince, S. A., & Poppel, M. N. M. van. (2021). Physical activity self-reports: Past or future? *British Journal of Sports Medicine*, *55*(16), 889–890. https://doi.org/10.1136/bjsports-2020-103595

Schmikat, M. (2025). *Messreaktivität bei der Erfassung körperlicher Aktivität: Eine Untersuchung von Selbstberichtsdaten im Rahmen des Projekts „Why does reactivity to physical activity measurement occur? Empirical test of potential mechanisms“* [Unpublished master’s thesis]. University of Vienna.

Shaw, M., & Flake, J. K. (n.d.). *Assumptions*. Introduction to Multilevel Modelling. https://www.learn-mlms.com/12-module-12.html

Shaw, M., Rights, J., Sterba, S., & Flake, J. (2025). *r2mlm: R-Squared Measures for Multilevel Models* (Version 0.3.8) [Software]. https://cran.r-project.org/web/packages/r2mlm/index.html

Silfee, V. J., Haughton, C. F., Jake-Schoffman, D. E., Lopez-Cepero, A., May, C. N., Sreedhara, M., Rosal, M. C., & Lemon, S. C. (2018). Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review. *Preventive Medicine Reports*, *11*, 74–80. https://doi.org/10.1016/j.pmedr.2018.05.003

Skender, S., Ose, J., Chang-Claude, J., Paskow, M., Brühmann, B., Siegel, E. M., Steindorf, K., & Ulrich, C. M. (2016). Accelerometry and physical activity questionnaires—A systematic review. *BMC Public Health*, *16*(1), Article 515. https://doi.org/10.1186/s12889-016-3172-0

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Strain, T., Flaxman, S., Guthold, R., Semenova, E., Cowan, M., Riley, L. M., Bull, F. C., Stevens, G. A., Raheem, R. A., Agoudavi, K., Anderssen, S. A., Alkhatib, W., Aly, E. A. H., Anjana, R. M., Bauman, A., Bovet, P., Moniz, T. B., Bulotaitė, G., Caixeta, R., … Zoma, L. R. (2024). National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: A pooled analysis of 507 population-based surveys with 5·7 million participants. *The Lancet Global Health*, *12*(8), e1232–e1243. https://doi.org/10.1016/S2214-109X(24)00150-5

Strain, T., Wijndaele, K., Garcia, L., Cowan, M., Guthold, R., Brage, S., & Bull, F. C. (2020). Levels of domain-specific physical activity at work, in the household, for travel and for leisure among 327 789 adults from 104 countries. *British Journal of Sports Medicine*, *54*(24), 1488–1497. https://doi.org/10.1136/bjsports-2020-102601

Strath, S. J., Kaminsky, L. A., Ainsworth, B. E., Ekelund, U., Freedson, P. S., Gary, R. A., Richardson, C. R., Smith, D. T., & Swartz, A. M. (2013). Guide to the assessment of physical activity: Clinical and research applications. *Circulation*, *128*(20), 2259–2279. https://doi.org/10.1161/01.cir.0000435708.67487.da

Tudor-Locke, C., & Bassett, D. R. (2004). How many steps/day are enough? *Sports Medicine*, *34*(1), 1–8. https://doi.org/10.2165/00007256-200434010-00001

Tudor-Locke, C., Bassett, D. R., Jr., Swartz, A. M., Strath, S. J., Parr, B. B., Reis, J. P., DuBose, K. D., & Ainsworth, B. E. (2004). A preliminary study of one year of pedometer self-monitoring. *Annals of Behavioral Medicine*, *28*(3), 158–162. https://doi.org/10.1207/s15324796abm2803_3

Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2009). Accelerometer-determined steps per day in US adults. *Medicine & Science in Sports & Exercise*, *41*(7), 1384–1391. https://doi.org/10.1249/MSS.0b013e318199885c

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*, *18*(2), 185–191. https://doi.org/10.1123/jpah.2020-0331

World Health Organization. (2018). *Global action plan on physical activity 2018-2030: More active people for a healthier world*. https://iris.who.int/bitstream/handle/10665/272722/9789241514187-eng.pdf

World Health Organization. (2020). *WHO guidelines on physical activity and sedentary behaviour*. https://iris.who.int/bitstream/handle/10665/336656/9789240015128-eng.pdf

List of Abbreviations

WHO	World Health Organisation
MET	metabolic equivalent
NHANES	National Health and Nutrition Examination Survey
LABS	University of Vienna Laboratory Administration for Behavioural Sciences

List of Figures

List of Tables

Appendix A: Additional Tables and Figures

Table A 1

Salience Manipulation Text

Salience condition	Original German text	English translation
low salience	Bei Ihrem nächsten Termin in zwei Wochen werden wir Sie bitten, nochmal einen Fragebogen auszufüllen. Dieser Fragebogen wird die gleichen Fragen enthalten wie der Fragebogen, den Sie heute ausgefüllt haben.	At your second appointment in two weeks, you will be asked to complete another survey. This survey will ask the same questions as the survey you completed today.
high salience	Bei Ihrem nächsten Termin in zwei Wochen werden wir Sie bitten, nochmal einen Fragebogen auszufüllen. In diesem Fragebogen werden Sie gebeten, Ihre körperliche Aktivität an jedem Tag zwischen dem heutigen Tag und dem nächsten Termin zu berichten. Indem Sie Ihre körperliche Aktivität in den nächsten zwei Wochen gewissenhaft beobachten, werden Sie dazu beitragen, wertvolle Erkenntnisse über Aktivitätsmuster im Alltag zu erlangen.	At your second appointment in two weeks, you will be asked to complete another survey. This survey will ask you to report on your physical activity for each day of the period between today and the next appointment. By carefully monitoring your physical activity in the next two weeks, you will help us to gain valuable insights into physical activity patterns in daily life.

Table A 2

Descriptive Characteristics of Participants Separately by Experience and Salience Condition

Variable	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	n/M	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*	*n/M*	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*
Number of participants	13	41.93%	18	58.07%			18	58.07%	13	41.93%	1.15 [c]	.462 [c]
of which experienced trackers							9	50%	4	30.77%		
of which novice trackers							9	50%	9	69.23%		
Age	25.69	11.2	20.83	1.82	4.5	.043*	21.78	5.3	24.39	10.1	1.03	.32
Gender [a]					1.87	.228					0.29	.689
female	8	61.54%	15	83.33%			14	77.78%	9	69.23%		
male	5	38.46%	3	16.67%			4	22.22%	4	30.77%		
other	0	0%	0	0%			0	0%	0	0%		
Highest level of education [a]					4.61	.38					2.96	.924

Variable	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	n/M	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*	*n/M*	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*
no school-leaving certificate	0	0%	0	0%			0	0%	0	0%		
Some high school, to 8th grade	1	7.69%	0	0%			0	0%	1	7.69%		
Some high school, no university entrance diploma	0	0%	0	0%			0	0%	0	0%		
High school graduate, university entrance diploma or the equivalent (Abitur/ Matura)	9	69.23%	16	88.89%			14	77.78%	11	84.62%		
Some college, no degree	1	7.69%	2	11.11%			2	11.11%	1	7.69%		

Variable	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	n/M	*%/SD*	*n/M*	*%/SD*	F/Chi^2	*p*	*n/M*	*%/SD*	*n/M*	*%/SD*	F/Chi^2	*p*
Trade/ technical/ vocational training	1	7.69%	0	0%			1	5.56%	0	0%		
Bachelor's degree	1	7.69%	0	0%			1	5.56%	0	0%		
Master's degree	0	0%	0	0%			0	0%	0	0%		
Doctoral degree	0	0%	0	0%			0	0%	0	0%		
Student [a, b]	11	84.62%	18	100%	2.96	.168	17	94.44%	12	92.31%	0.06	1
Household net income [a]					7.83	.494					10.29	.167
less than 150€	1	7.69%	4	22.22%			1	5.56%	4	30.77%		
150€ to 300€	1	7.69%	1	5.56%			1	5.56%	1	7.69%		
300€ to 500€	0	0%	3	16.67%			2	11.11%	1	7.69%		
500€ to 1,000€	5	38.46%	3	16.67%			7	38.89%	1	7.69%		
1,000€ to 1,500€	1	7.69%	1	5.56%			1	5.56%	1	7.69%		

Variable	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	n/M	*%/SD*	*n/M*	*%/SD*	F/Chi^2	*p*	*n/M*	*%/SD*	*n/M*	*%/SD*	F/Chi^2	*p*
1,500€ to 2,000€	1	7.69%	1	5.56%			1	5.56%	1	7.69%		
2,000€ to 2,500€	1	7.69%	0	0%			0	0%	1	7.69%		
2,500€ to 3,000€	1	7.69%	4	22.22%			2	11.11%	3	23.08%		
3,000€ to 5,000€	2	15.39%	1	5.56%			3	16.67%	0	0%		
5,000€ to 10,000€	0	0%	0	0%			0	0%	0	0%		
more than 10,000€	0	0%	0	0%			0	0%	0	0%		
BMI	24.23	3.15	22.96	2.81	2.04	.165	23.16	3.12	23.94	2.81	0.59	.451
presence of respiratory disease in the past 14 days [a, b]	1	7.69%	0	0%	1.43	.419	1	5.56%	0	0%	0.75	1

Variable	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	n/M	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*	*n/M*	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*
presence of injury in the past 14 days [a, b]	2	15.39%	0	0%	2.96	.168	1	5.56%	1	7.69%	0.06	1
presence of stressful life events in the past 14 days [a, b]	5	38.46%	6	33.3%	0.09	1	7	38.89%	4	30.77%	0.22	.718
For experienced physical activity trackers:												
Duration of tracker usage in months	33.58	12.92					31.63	15	37.5	7.55	0.9	.388
Self-reported consistency of tracker usage in general (1 = *not at*	4.08	1.16					3.75	1.28	4.75	0.5	1.93	.083

Variable	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	n/M	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*	*n/M*	*%/SD*	*n/M*	*%/SD*	*F/Chi*^{2}	*p*
all consistent to 5 = *very consistent*)												

Note. T-test were conducted with Welch's correction.

[a] Reported p-values according to Fisher's exact test because of expected frequencies <5 in more than 20% of cells. [b] Absolute and relative frequencies of participants answering "yes". [c] Chi^2-test values relate to the 2 (experience) x 2 (salience) crosstable.

* p-values are significant at α = .05.

Table A 3

Accelerometer Wear Time Separately by Tracking Experience and Salience Condition

Day in study period	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	M	*SD*	*M*	*SD*	*F*	*p*	*M*	*SD*	*M*	*SD*	*F*	*p*
Over days 1-15	11:29:19	04:07:52	11:51:46	04:00:33	0.55	.457	11:29:49	04:16:18	11:59:43	03:44:25	1.71	.192
Over days 1-14	12:05:59	03:26:44	12:30:20	03:13:57	0.96	.327	12:07:40	03:33:15	12:37:23	02:57:59	2.36	.126

Day in study period	Experienced Trackers		Novice Trackers		Tracking experience comparison		High Salience		Low Salience		Salience condition comparison	
	M	*SD*	*M*	*SD*	*F*	*p*	*M*	*SD*	*M*	*SD*	*F*	*p*
Day 1	10:09:00	02:33:28	09:42:53	02:02:13	0.03	.861	10:32:33	02:32:17	09:00:14	01:23:14	3.66	.066
Day 2	13:05:28	01:42:53	13:51:53	03:16:59	0.29	.598	12:58:53	03:11:00	14:18:51	01:45:51	1.82	.188
Day 3	12:13:14	02:29:59	12:45:30	03:04:51	0.1	.759	12:03:40	02:37:55	13:11:09	03:02:39	1.13	.297
Day 4	12:06:37	02:45:01	11:21:43	03:15:00	0.54	.471	11:36:20	03:33:55	11:46:23	02:12:10	0.02	.879
Day 5	12:44:09	02:59:32	12:33:17	03:55:29	0.24	.631	11:43:20	04:03:30	13:53:18	02:07:27	2.99	.095
Day 6	12:39:00	02:26:54	12:18:30	02:53:59	0.02	.894	12:54:23	02:59:38	11:49:18	02:07:51	1.16	.29
Day 7	12:46:18	04:41:41	12:49:40	02:51:36	0	.987	12:37:57	04:03:18	13:02:32	03:11:54	0.09	.771
Day 8	12:19:51	03:25:37	13:56:23	02:15:33	2.21	.149	13:03:00	03:23:24	13:33:46	02:02:19	0.25	.625
Day 9 [a]	10:43:37	03:45:21	12:56:03	02:13:11	0.13	.722	10:45:43	03:23:35	13:44:05	01:30:42	9.28	.005*
Day 10	12:42:18	04:11:32	13:17:40	03:54:09	0.03	.865	12:17:23	03:56:36	14:05:46	03:54:34	1.58	.22
Day 11	11:20:46	02:49:40	13:12:00	04:21:25	1.42	.245	12:00:03	04:33:15	13:00:23	02:39:06	0.5	.486
Day 12	11:32:51	03:29:32	12:56:27	02:31:16	1.6	.217	12:17:23	02:39:55	12:26:55	03:30:47	0.02	.886
Day 13	11:33:09	05:07:14	11:31:07	03:48:00	0.02	.881	12:07:37	04:21:27	10:42:37	04:18:12	0.78	.386
Day 14	13:27:28	04:05:28	11:51:40	02:17:33	1.6	.217	12:48:57	03:50:35	12:08:09	02:08:17	0.33	.57
Day 15	02:55:55	03:26:23	02:51:47	02:51:26	0.02	.882	02:40:00	03:42:34	03:12:14	01:55:49	0.21	.649

Note. Accelerometer wear time in hh:mm:ss format.

[a] ANOVAs conducted with White's correction for heteroscedasticity due to significant Levene's tests.

* p-values are significant at $\alpha = .05$.

Table A 4

Steps per Day Separately by Tracking Experience and Salience Condition

Day in study period	Experienced Trackers		Novice Trackers		High Salience		Low Salience	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Over days 1-15	8,181	5,044	6,925	3,519	7,773	4,495	7,019	3,912
Over days 1-14	8,553	4,982	7,142	3,445	8,065	4,424	7,280	3,871
Day 1	8,433	5,747	6,669	2,999	8,301	5,122	6,174	2,775
Day 2	9,482	4,937	7,562	4,783	7,465	3,877	9,616	5,907
Day 3	8,829	3,025	8,550	3,643	8,031	3,579	9,548	2,900
Day 4	8,284	3,747	7,042	2,833	7,727	3,168	7,335	3,474
Day 5	8,122	3,810	6,956	2,717	7,429	3,154	7,503	3,453
Day 6	9,885	5,008	6,106	3,676	8,489	4,878	6,584	4,149
Day 7	10,322	9,309	6,368	3,504	9,279	7,854	6,292	4,570
Day 8	8,474	5,795	8,011	2,453	8,671	4,593	7,560	3,403
Day 9	7,589	3,963	6,732	2,996	6,963	3,006	7,268	4,003
Day 10	6,783	3,673	7,632	2,847	7,309	3,150	7,231	3,371
Day 11	8,550	4,739	7,568	4,792	8,561	4,900	7,176	4,514
Day 12	7,325	4,356	7,102	3,013	7,910	4,072	6,206	2,559
Day 13	8,523	4,549	6,157	3,714	7,683	4,255	6,409	4,133
Day 14	9,141	5,441	7,520	3,571	9,052	4,863	7,019	3,637
Day 15	2,972	2,338	3,517	2,930	3,199	2,895	3,363	2,426

Table A 5

Model Parameters for Hypothesis 1 (Including All Days)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	7,995.96	628.45	76	12.72	< .001*
Time	-101.84	52.32	431.33	-1.95	.052
Tracking experience	1,979.85	953.29	71.07	2.08	.041*
Interaction of Time *	-103.27	80.29	431.18	-1.29	.199
Tracking Experience					
Weekend/-day	-1,346.65	386.51	431.09	-3.48	< .001*
Covariate					
Random effects					
Intercept	3,585,260	1,893			
Residual variance	13,476,676	3,671			
Number of Parameters	7				
-2LL	8,945.2				
AIC	8,959.2				
BIC	8,988.1				

* p-values are significant at α = .05.

Table A 6

Model Parameters for Hypothesis 2 (Including All Days)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	8,722.11	745.90	70.91	11.69	< .001*
Time	-193.49	60.89	431.05	-3.18	.002*
Salience	158.42	971	68.78	0.16	.871
Interaction of Time *	83.87	80.41	431.16	1.04	.298
Salience					
Weekend/-day	-1,307.47	387.12	431.07	-3.38	.001*
Covariate					
Random effects					
Intercept	3,833,005	1,958			
Residual variance	13,494,788	3,674			
Number of Parameters	7				
-2LL	8,947.4				
AIC	8,961.4				
BIC	8,990.4				

* p-values are significant at α = .05.

Table A 7

Model Parameters for Hypothesis 3 (Including All Days)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	8,100.34	793.77	70.10	10.21	< .001*
Time	-115.80	70.33	458.49	-1.65	.1
Tracking experience	2,024.38	975.47	68.44	2.08	.042*
Interaction of Time *	-252.48	116.58	388.63	-2.166	.031*
Tracking Experience					
Salience	-232.47	975.82	68.53	-0.24	.812
Interaction of Time *	28.78	97.08	454.56	0.3	.767
Salience					
Interaction of Time *	207.47	136.08	209.22	1.53	.129
Tracking Experience *					
Salience					
Weekend/-day	-1,311.63	384.78	430.45	-3.41	.001*
Covariate					
Random effects					
Intercept	3,674,987	1,917			
Residual variance	13,319,635	3,650			
Number of Parameters	10				
-2LL	8,940.7				
AIC	8,960.7				
BIC	9,002.0				

* p-values are significant at $\alpha = .05$.

Table A 8

Model Parameters for Hypothesis 1 (Excluding Day 1 and Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	8,041.28	665.51	70.03	12.08	< .001*
Time	-49.50	60.93	371.06	-0.81	.417
Tracking experience	1,762.49	1,006.92	64.85	1.75	.085
Interaction of Time *	-64.05	94.02	371.04	-0.68	.496
Tracking Experience					
Weekend/-day	-1,820.72	377.57	371.04	-4.82	< .001*
Covariate					
Random effects					
Intercept	4,313,339	2,077			
Residual variance	12,108,253	3,480			
Number of Parameters	7				
-2LL	7,750.7				
AIC	7,764.7				
BIC	7,792.7				

* p-values are significant at α = .05.

Table A 9

Model Parameters for Hypothesis 2 (Excluding Day 1 and Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	9,041.60	791.19	64.5	11.43	< .001*
Time	-189.58	71.16	371.03	-2.66	.008*
Salience	-494.58	1,028.62	62.24	-0.48	.632
Interaction of Time * Salience	196.29	93.71	371.04	2.1	.037*
Weekend/-day	-1,751.44	376.32	371.04	-4.65	< .001*
Covariate					
Random effects					
Intercept	4,671,873	2,161			
Residual variance	11,981,607	3,461			
Number of Parameters	7				
-2LL	7,748.8				
AIC	7,762.8				
BIC	7,790.8				

* p-values are significant at α = .05.

Table A 10

Model Parameters for Hypothesis 3 (Excluding Day 1 and Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	8,448.51	838.61	64.24	10.07	< .001*
Time	-104.48	82.46	395.32	-1.27	.206
Tracking experience	1,926.76	1,028.97	62.34	1.87	.066
Interaction of Time *	-276.58	137.99	365.12	-2	.046*
Tracking Experience					
Salience	-866.91	1,029.62	62.49	-0.84	.403
Interaction of Time *	111.20	113.21	400.83	0.98	.327
Salience					
Interaction of Time *	276.91	162.41	230.61	1.71	.09
Tracking Experience *					
Salience					
Weekend/-day	-1,750.65	374.04	370.59	-4.68	< .001*
Covariate					
Random effects					
Intercept	4,433,207	2,106			
Residual variance	11,823,102	3,438			
Number of Parameters	10				
-2LL	7,742.4				
AIC	7,762.4				
BIC	7,802.4				

* p-values are significant at α = .05.

Table A 11

Set of R^2s for the Main Hypothesis 1 Model

	total	within	between
f_1	.0341	.0454	NA
f_2	.0272	NA	.1091
v_1	.0000	.0000	NA
v_2	.0000	NA	.0000
m	.2222	NA	.8909
f	.0613	NA	NA
fv	.0613	.0454	.1091
fvm	.2835	NA	NA

Note. The full decomposition of variance for models including unbalanced Level-1 predictors which are not person-mean-centred differentiates between v_1, reflecting the within-person-varying portion of Level-1 predictors via random slope variation, and v_2, reflecting the between-person-varying portion of Level-1 predictors via random slope variation. For models with completely balanced and/or person-mean-centred Level-1 predictors, there is no between-person variation of Level-1 predictors and thus no v_2, so v_1 is termed simply v, reflecting Level-1 predictors via random slope variation. This differentiation between v_1 and v_2 was not mentioned in the main text because the definitions of sources of explained variance were limited to those derived under assumptions of person-mean-centring. As the R function for the decomposition of variance for longitudinal models outputs both parameters, full results are, however, reported here. Since the present model does not include random slopes associated with any of the predictors, both v_1 and v_2 equalled 0 in any case here.

Table A 12

Set of R^2s for the Main Hypothesis 2 Model

	total	within	between
f_1	.0363	.0484	NA
f_2	.0086	NA	.0343
v_1	.0000	.0000	NA
v_2	.0000	NA	.0000
m	.241	NA	.9657
f	.0449	NA	NA
fv	.0449	.0484	.0343
fvm	.2859	NA	NA

Note. See Note to Table A11.

Table A 13

Set of R^2s for the Main Hypothesis 3 Model

	total	within	between
f_1	.0409	.0557	NA
f_2	.0398	NA	.1493
v_1	.0000	.0000	NA
v_2	.0000	NA	.0000
m	.2266	NA	.8507
f	.0806	NA	NA
fv	.0806	.0557	.1493
fvm	.3072	NA	NA

Note. See Note to Table A11.

Table A 14

Model Parameters for the Unconditional Model of Change With Only Study Day and the Weekend/Weekday Covariate (Excluding Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	8,442.79	509.67	69.54	16.57	< .001*
Time	-35.18	42.46	402.03	-0.83	.408
Weekend/-day	-1,667.44	378.91	402.03	-4.4	< .001*
Covariate					
Random effects					
Intercept	4,414,939	2,101			
Residual variance	12,704,132	3,564			
Number of Parameters	5				
-2LL	8,366.4				
AIC	8,376.4				
BIC	8,396.7				

* p-values are significant at α = .05.

Table A 15

Set of R^2s for the Unconditional Model of Change With Only Study Day and the Weekend/Weekday Covariate (Excluding Day 15)

	total	within	between
f_1	.0333	.0443	NA
f_2	.0000	NA	.0000
v_1	.0000	.0000	NA
v_2	.0000	NA	.0000
m	.2493	NA	1
f	.0333	NA	NA
fv	.0333	.0443	.0000
fvm	.2826	NA	NA

Note. See Note to Table A11.

Table A 16

Model Parameters for the Reduced Model for Hypothesis 3 Without the Three-way Interaction (Excluding Day 15)

Effect	*b*	*SE*	*df*	*t*	*p*
Fixed effects					
Intercept	7,846.69	804.32	69.01	9.76	< .001*
Time	-76.60	70.74	402.02	-1.08	.28
Tracking experience	1,863.91	990.13	67.82	1.88	.064
Interaction of Time *	-86.41	87.41	402.03	-0.99	.324
Tracking Experience					
Salience	-334.22	990.63	67.95	-0.34	.737
Interaction of Time *	133.92	87.54	402.03	1.53	.127
Salience					
Weekend/-day	-1,642.80	378.51	402.03	-4.34	< .001*
Covariate					
Random effects					
Intercept	3,871,740	1,968			
Residual variance	12,615,188	3,552			
Number of Parameters	9				
-2LL	8,360.1				
AIC	8,378.1				
BIC	8,414.8				

* p-values are significant at $\alpha = .05$.

Table A 17

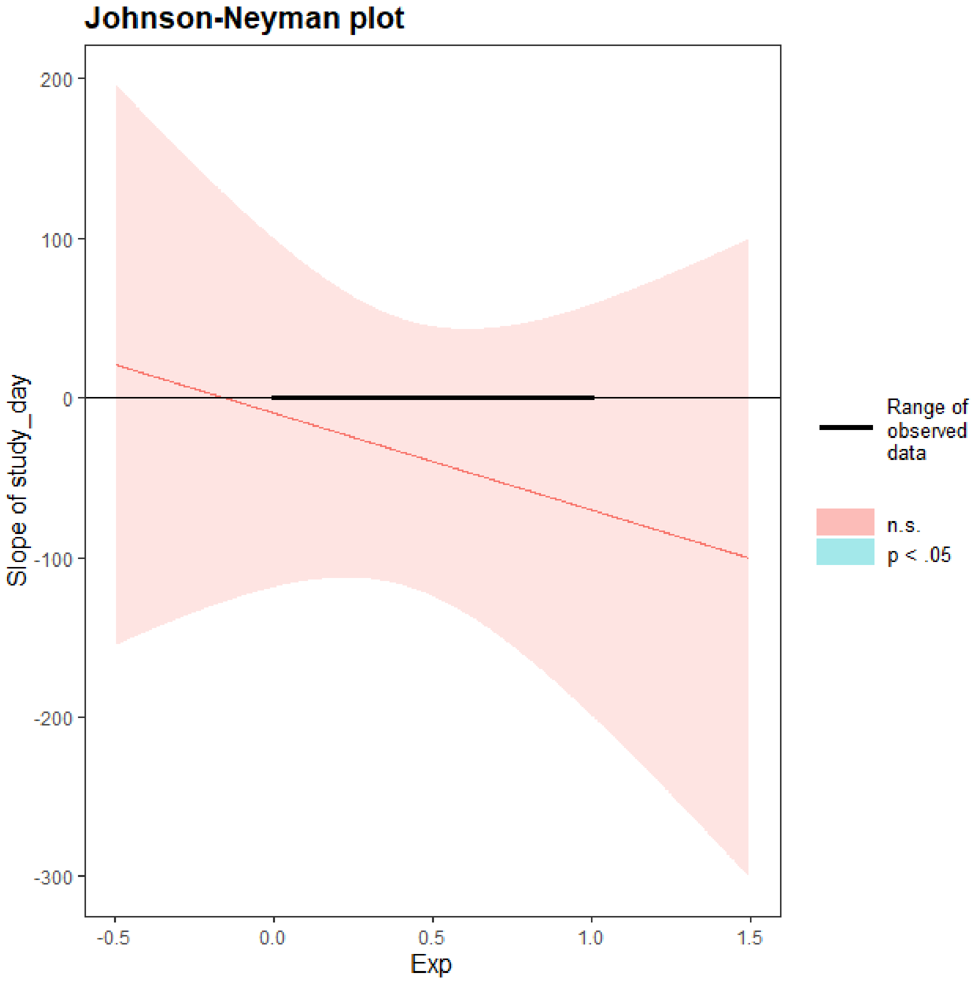
Set of R^2s for the Reduced Model for Hypothesis 3 Without the Three-way Interaction (Excluding Day 15)

	total	within	between
f_1	.0379	.0505	NA
f_2	.031	NA	.1243
v_1	.0000	.0000	NA
v_2	.0000	NA	.0000
m	.2186	NA	.8757
f	.069	NA	NA
fv	.069	.0505	.1243
fvm	.2876	NA	NA

Note. See Note to Table A11.

Figure A 1

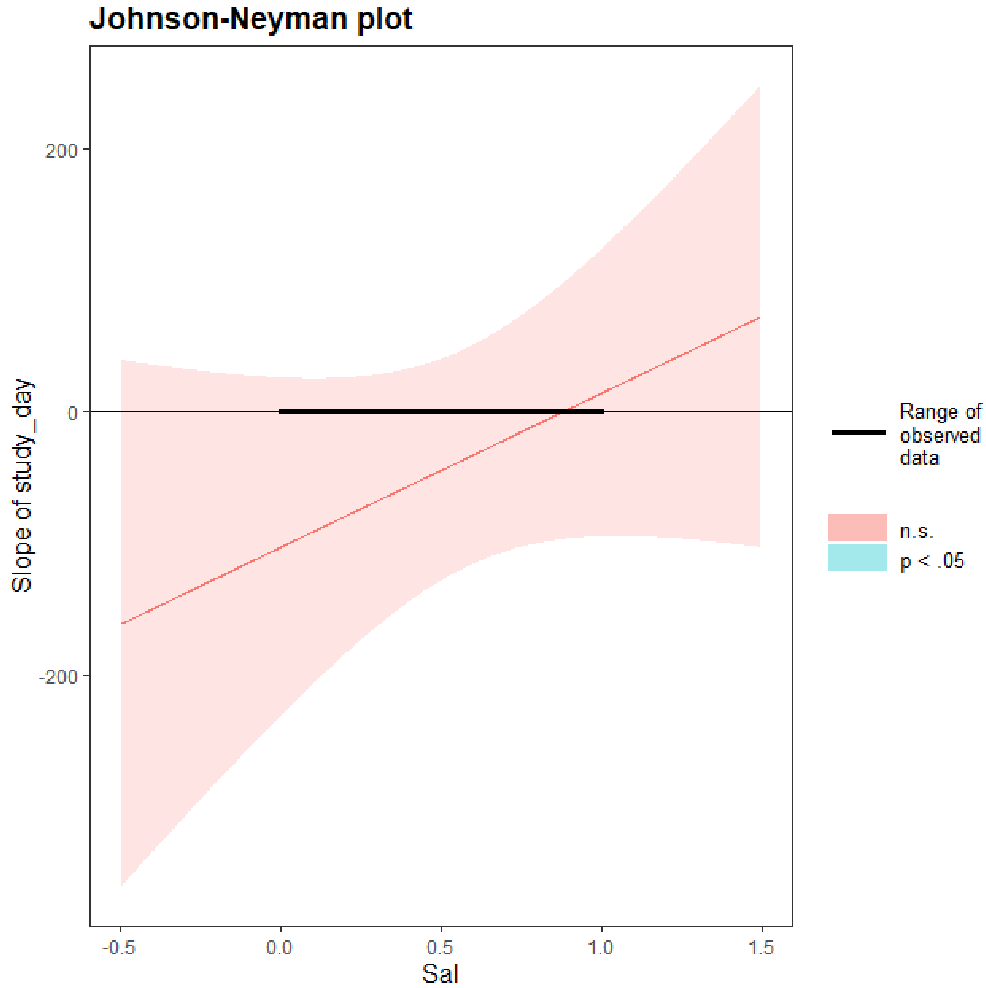
Johnson-Neyman Plot Describing the Interaction Between Study Day and Tracking Experience in the Main Hypothesis 1 Model



Note. Only 0 (novice trackers) and 1 (experienced trackers) were observed values for tracking experience; all other values are hypothetical.

Figure A 2

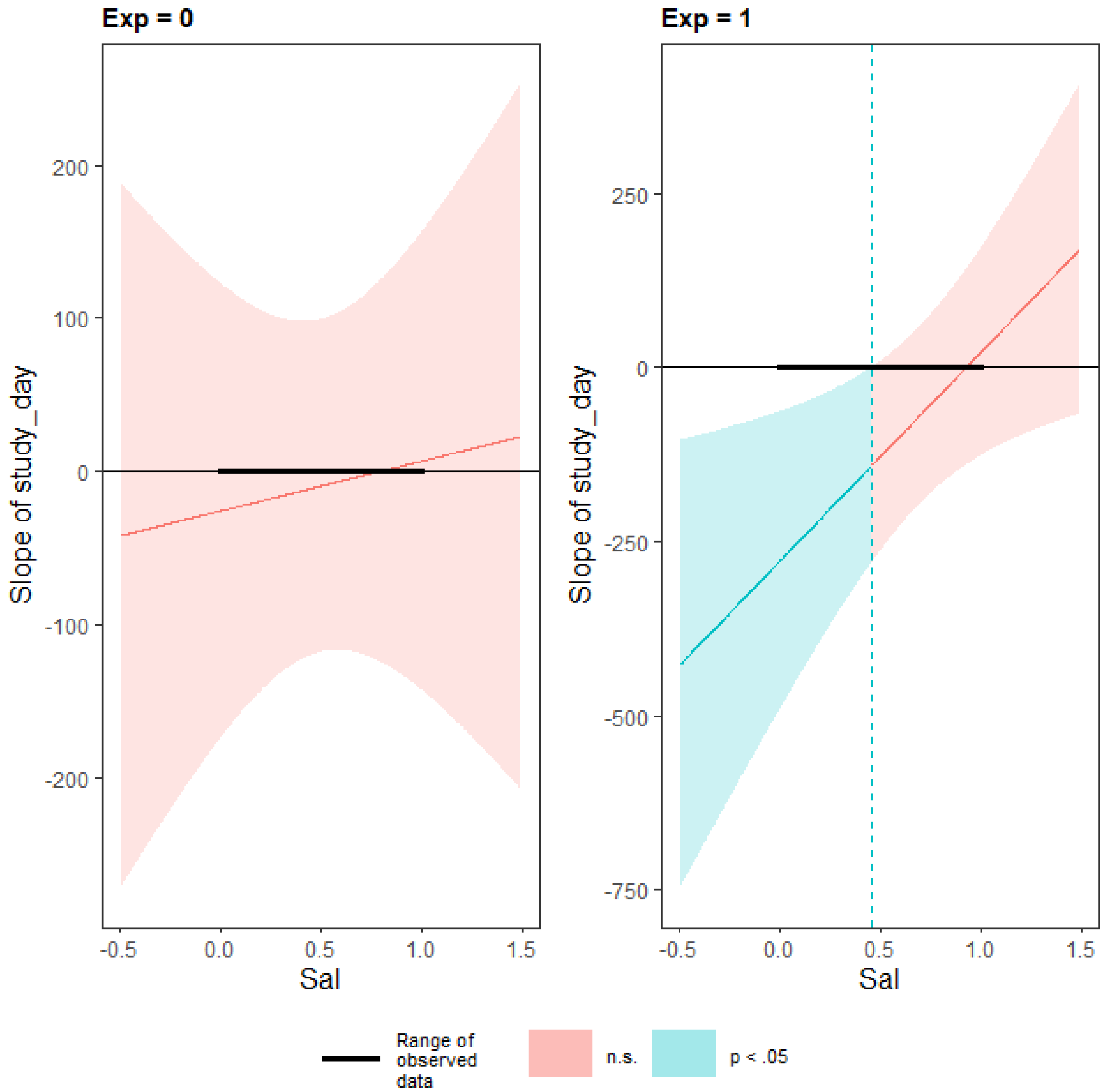
Johnson-Neyman Plot Describing the Interaction Between Study Day and Salience Condition in the Main Hypothesis 2 Model



Note. Only 0 (low salience condition) and 1 (high salience condition) were observed values for salience condition; all other values are hypothetical.

Figure A 3

Johnson-Neyman Plot Describing the Interaction Between Study Day, Salience Condition, and Tracking Experience in the Main Hypothesis 3 Model



Note. The figure on the left-hand side depicts the relationship between study day and salience in novice trackers (Exp = 0). The figure on the right-hand side depicts the relationship between study day and salience in experienced trackers (Exp = 1). Only 0 (low salience condition) and 1 (high salience condition) were observed values for salience condition; all other values are hypothetical. In the right-hand figure, the blue dashed line indicates the boundary of the region of significance at the hypothetical value of 0.46 salience.

Figure A 4

Visualisation of the R^2 Set for the Main Hypothesis 1 Model

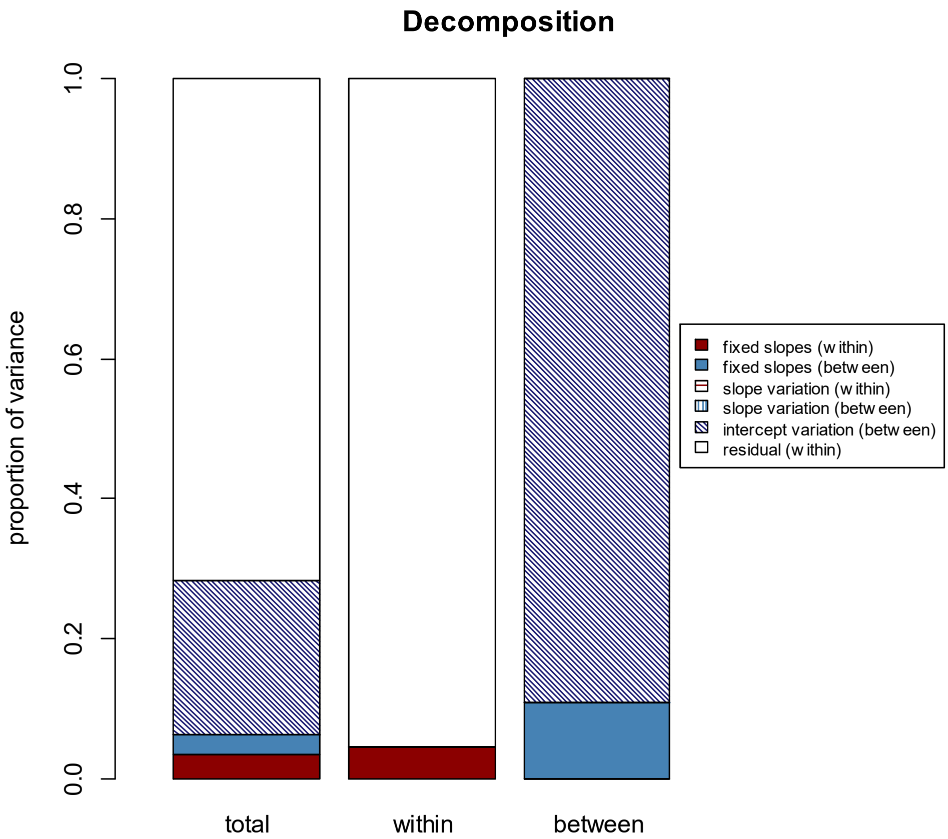


Figure A 5

Visualisation of the R^2 Set for the Main Hypothesis 2 Model

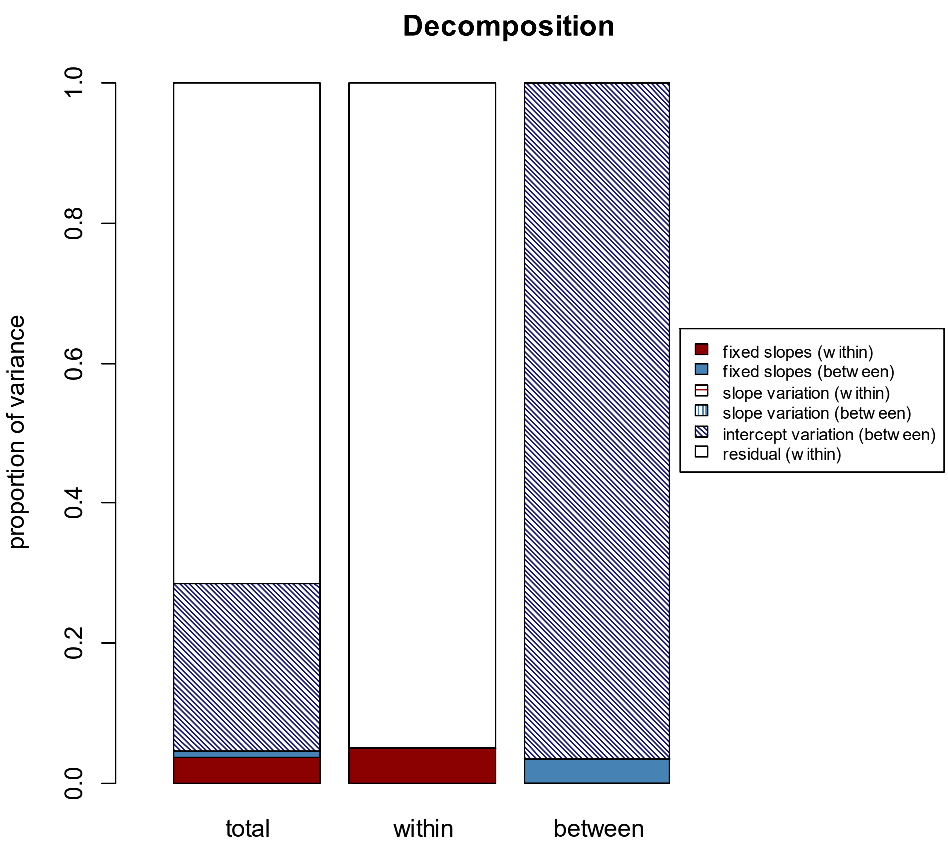
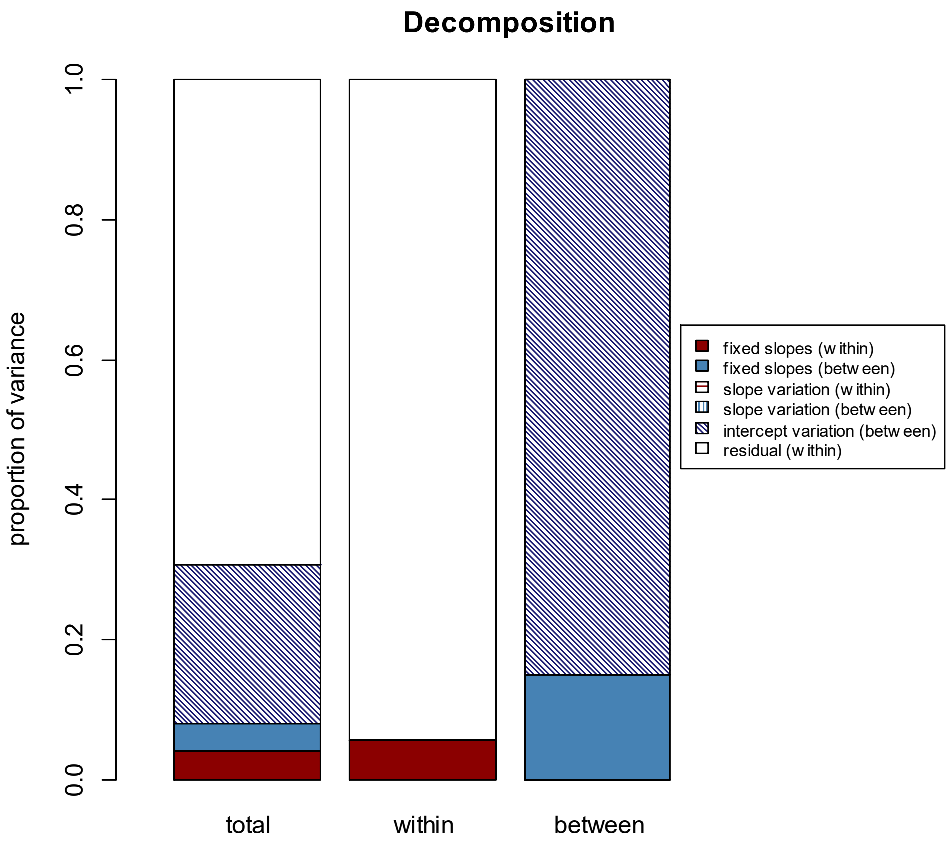


Figure A 6

Visualisation of the R^2 Set for the Main Hypothesis 3 Model



Appendix B: Declaration on the Use of Artificial Intelligence (AI)

Erklärung zur Verwendung von Künstlicher Intelligenz (KI)

Hiermit erkläre ich, dass ich die folgenden KI-Hilfsmittel zur Erstellung meiner Masterarbeit mit dem Titel „Measurement Reactivity in Experienced versus Novice Physical Activity Trackers" verwendet habe. Ich versichere, dass die Kennzeichnung der verwendeten KI-Hilfsmittel vollständig ist, inkl. Verwendungszweck, Produktname, als auch ggfs. Prompts und Outputs des KI-Hilfsmittels. Ich verantworte die Auswahl und die Übernahme sämtlicher Ergebnisse der von mir verwendeten KI-Hilfsmittel vollumfänglich selbst.

Verwendungszweck	**Produktname des KI-Hilfsmittels**
Verstehen, Erstellen und Überarbeiten von Teilen der R Syntax für die statistischen Analysen	StatsBot (MethodsHub)

Alternativ:

☐ Ich bestätige, dass ich zur Erstellung meiner Masterarbeit keine KI-Hilfsmittel verwendet habe.