Rethinking Subjective Metrics in Self-Tracking Systems for Running

Master Thesis
Interaction Technology
University of Twente

Wouter Sebastiaan Harmen van Mensvoort
Supervised by
Dennis Reidsma
Armağan Karahanoğlu

26-06-2025

Abstract

Self-tracking technology in sports primarily focuses on performance enhancement and goal completion. While this approach can improve performance, it may also negatively impact how athletes experience their sport. We highlight the lack of subjective metrics in self-tracking technology and their potential to enhance the Sport-Data Experience (SDX). To address this, we explored opportunities to enhance the SDX and identified interoceptive awareness as an area to focus on. Next, we selected a suitable subjective metric designed to help increase this awareness and used it to develop an intervention, which we implemented within a smartwatch-based self-tracking application for use during running. Finally, we evaluated this intervention using a mixed-methods evaluation involving 21 participants in which we compared a baseline group with the intervention group. Results show a statistically significant increase in interoceptive awareness in the intervention group ($r = .44$), supported by qualitative think-aloud and interview data. These findings demonstrate that subjective metrics can meaningfully improve runners' engagement with their bodily sensations and suggest design opportunities for more reflective and personalised self-tracking tools.

Acknowledgement

As a student coming from HBO rather than a university bachelor’s program, I found it quite challenging at times to shift into a more academic, research focused mindset while working on this Master’s thesis. Luckily for me, I had great supervision. I want to sincerely thank my supervisors, Armağan Karahanoğlu and Dennis Reidsma, for the weekly meetings, detailed feedback, and tough questions that pushed me to think more critically and continuously improve the quality of this work. I often left those meetings with more questions than I came in with, but they were always helpful for me to move forward. Thanks also to my family, friends, and fellow students who participated in the study. I genuinely listened with interest to your thoughts and experiences regarding the intervention. Those conversations made the research feel more real, and for me, they were the most enjoyable part of the entire process. Now my time at the University of Twente has almost come to an end and I look forward to putting my gained knowledge to work.

Table of Contents

1. Introduction

Over the past year, the global fitness tracker market has experienced significant growth, reaching a valuation of USD 62.03 billion in 2024[1]. This market is expected to expand further at an annual rate of 21.3% over the next eight years, reaching a total value of USD 290.85 billion by 2032. Fitness trackers enable continuous monitoring of metrics such as steps taken, heart rate and sleep patterns (Feng et al., 2020). In sport, self-tracking is often used to improve performance, achieve goals and prevent injuries (Karahanoğlu et al., 2021). Despite the potential of self-tracking devices to increase well-being, there are news reports on how some users of self-tracking technology do not achieve the intended effects of the devices [2, 3]. While the devices are designed to track and promote health and physical activity, they can also have unintended demotivating effects like feeling discouraged when unable to meet default goals, such as achieving 10,000 steps per day.

Loerakker et al. (2023) point out that most fitness trackers are primarily focused on performance improvement and goal achievement. This heavy emphasis on numerical data can create a very performance-driven experience. While goal setting can improve athletic performance, it also presents challenges. Negative feedback can undermine athletes' self-confidence, and failing to reach goals can damage wellbeing, encourage negative thought patterns, and reduce motivation. Similarly, Tholander & Nylander (2015) highlight how the constant stream of performance data can create pressure, potentially reducing the enjoyment of sport. They emphasise the dual nature of sports technology, where athletes assess their training through both measured performance and perceived experience. They emphasise that future technology should better integrate a balance between these two. This viewpoint is also shared by (Saw et al., 2015), who suggest that subjective self-reported measures offer valuable insights into athlete well-being, which could complement performance data in creating a more holistic understanding of an athlete’s experience.

From this perspective, it becomes clear that sports technology plays a crucial role in shaping how athletes experience their sport. This interaction between athletes and data is captured by Postma et al. (2024) in the concept of Sport-Data Experience (SDX). They define SDX as “the subjective and multifaceted way athletes interact with, perceive, and are affected by sport data in the context of their training, performance, and overall engagement in their sport.” SDX goes beyond just analysing numbers and statistics from a smartwatch or app. It focuses on how athletes **perceive, interpret and react to data** they encounter in their training routines. This includes using data to make sense of one’s own training practice, make informed decisions, adjust training intensity, set personal goals and engage with others for support and encouragement. While tracking progress can boost motivation and confidence when improvements are visible, it can also lead to frustration when expectations are not met.

In a preceding literature study, we identified several directions that have potential to improve the SDX of interactive system used by athletes. These include goal re-evaluation, qualitative goals, derived and subjective metrics, data sensemaking, gamification, and flow. Among these, subjective metrics stand out as a way to capture emotional and experiential aspects of performance that are often overlooked by purely quantitative data. Therefore, this thesis focuses on **subjective metrics** as a promising approach to enhance the Sport-Data Experience.

Subjective metrics can capture the personal dimensions of athletic performance, offering insights beyond traditional objective measures like speed or distance (Tholander & Nylander, 2015). They are not measured by any sensors but are instead manually entered by the athlete. While objective

[1] https://www.fortunebusinessinsights.com/fitness-tracker-market-103358
[2] https://edition.cnn.com/2016/09/01/health/dark-side-of-fitness-trackers/index.html
[3] https://www.techradar.com/news/do-fitness-trackers-really-help-with-motivation

data provides measurable results, it often neglects the physical sensations and emotions that make athletic experiences meaningful. Subjective metrics can capture athletes' lived experiences, such as exhaustion, discomfort, or accomplishment, offering a more holistic perspective that moves beyond a solely performance-oriented view of sports (Rapp & Tirabeni, 2018). Translating subjective sports experiences into meaningful insights is seen as one of the grand challenges in the field of SportHCI (Elvitigala et al., 2024). Furthermore, research by Karahanoğlu et al. (2024) also stresses the importance of subjective experience in training load management (TLM) and how insight into these can prevent injury and overtraining.

The importance of subjective metrics is highlighted in literature but there is a gap in the implementation of these metrics in real-word applications. For this reason, in this thesis we aim to develop and implement subjective metrics into self-tracking technology used by runners. The goal is to develop a metric that provides valuable insights for athletes and therefore improving their SDX. This brings forward the following research question:

How can a subjective metric be effectively integrated into self-tracking technology for running such that it enhances the experience of the athlete?

Addressing the main research question involves determining a suitable subjective metric, designing a practical way to implement it in runners' training, and evaluating its impact on their experience. This led to three sub-questions, each addressed in Chapters 2, 3, and 4 respectively:

1. **What opportunities exist for integrating a subjective metric into self-tracking technology to influence the runner's experience?**
2. **How to implement a subjective metric such that it is likely to have impact on interoceptive awareness?**
3. **What is the effect of the designed subjective metric on the runner's experience?**

In the first sub-question, we explored opportunities for implementing subjective metrics by examining their current use in popular running apps and reviewing relevant literature. Here we decided to focus on implementing a conspective (during-activity interaction; Postma et al., 2024) subjective metric with the goal of enhancing the athlete's awareness of their bodily sensations. To address the second question, we designed the subjective metric and implemented it into the self-tracking tech of runners. Once the metric was conceptualised, we moved to the final question in which we tested the newly formed metric, thus using it to create an intervention and evaluating its impact on the runner's experience. This study will support other researchers in the development of self-tracking technology that optimally balances performance enhancement and experience in sports, providing insights into effective implementations and how they impact athletes' experiences.

2. Exploring Design Opportunities

In this chapter, we explore the possibilities of integrating subjective metrics into self-tracking technology for running to improve the SDX. As mentioned earlier, SDX is influenced by various factors (Postma et al., 2024). Instead of aiming for a broad objective like enhancing the overall experience, our goal is to identify a specific aspect of the experience to focus on, ensuring a clear design direction. This also requires selecting a suitable and meaningful subjective metric. We start by analysing existing running apps to see the extent to which (and how) they already embed subjective metrics. Next, we turn to the literature to explore popular subjective metrics and their relevance in the context of sports training, as well as to examine related work in the field. Finally, we identify two possible directions in which these subjective metrics could improve aspects of SDX, and decide upon the final focus of this project.

2.1 Subjective Metrics in Running Tracking

Janssen et al. (2017) highlight that recreational runners frequently rely on physical devices like smartphones and wearable technology for self-tracking, with 85% of runners using at least one of these two. Since we do not have access to all the different smartwatches, and each smartwatch brand has its own designated application, our focus is instead on analysing the most popular run-tracking applications available in the Google Play Store. We analyse their implementations of subjective metrics, specifically examining the non-premium versions of Adidas Running, Runkeeper Asics, Nike Run Club, MapMyRun, Runna, Strava and Maprunner. Additionally, we include Training Peaks and Vortza due to personal familiarity and their advanced analysis features, as well as Garmin and Polar Flow, which are widely used platforms associated with dedicated sports wearables.

Each of the apps was analysed using the following questions:

- Which subjective metrics are implemented?
- How does the app collect this information?
- At what point does the app request this information?
- Does the app provide guidance on how these metrics can lead to insights?
- Is the subjective metric shown in relation to other data?

The result of the analysis can be found in Table 1 down below.

Table 1

Result analysis subjective metrics in running apps

App	Subjective metrics	Scale	Method	Timing	Guidance for insight	Integration with other data
Adidas Running	How did it feel? Notes	(1-5) Injured - Fantastic	Smilies with description	After workout	No	No
Runkeeper Asics	How did this run feel? Notes	(1-5) Painful to Great	Smilies with description	After workout	No	No
Nike Run Club	Exertion levels. -------- Notes	(1-10) Minimum to	Slider with description	After workout	No	No

		maximum Effort				
MapMyRun	Notes	-	-	After workout	No	No
Runna	How did you like this workout? -------- Notes	(1-2) Good - Bad	Thumbs up/down	After workout	No	No
Strava	Workout effort. -------- Notes	(1-10) Minimum to Maximum effort	Slider with description	After workout	Yes	Used to calculate Fitness & Freshness
Map runner	Training feeling. -------- Notes	(1-5) Easy - Exhausting	Smilies with description	After workout	No	No
Training Peaks	How did you feel? -------- Notes	(1-10) Minimum to maximum exertion	Smilies and slider with description	After workout	Yes	No
Vortza	Rate of Perceived Exertion -------- Notes	(1-10) Minimum to maximum effort	Slider with description	After workout	Yes	Used for calculating fitness and load balance
Garmin Connect	How did you feel? -------- Perceived Effort -------- Notes	(1-5) Very weak – Very Strong -------- (1 - 10) Perceived Effort	Smilies and slider with description	After workout	No	No
Polar Flow	How do you feel? -------- Notes	(1-5)	Smilies	After workout	No	No

All apps provided the functionality of adding a note to the workout in text format after it was recorded. Most apps also implemented a scale that could be used to record a subjective aspect after the workout. This was implemented differently among the various apps. Most of the apps featured a set of five emojis, ranging from a very unhappy face to a very happy one, with each emoji accompanied by a brief description. The implementations of these scales look similar to the Wong-Baker faces pain rating scale (WBFPS), often used in pain assessment (Khatri & Kalra, 2012). The sad emojis are supposed to indicate negative experience, while the happy ones denote a positive one. Figure 1 shows how Runkeeper Asics requests the subjective metric from the user. In this example, the description corresponding to the emoji is only visible after the activity is saved.

Figure 1
Subjective metric in Runkeeper Asics

Adidas, Runkeeper and Garmin Connect seemed to focus on capturing the overall feeling of the workout, with their descriptions ranging from injured/painful/very weak to fantastic/great/very strong. Also, Runna tried to capture feeling but simplifies the scale to just two options: thumbs up or thumbs down. Polar flow did also ask about feeling but does not add any description to the emojis. Nike Run Club, Strava, Vortza and Garmin Connect seemed to capture the rate of perceived exertion (RPE) in an almost identical way, all with slider ranging from 1-10, and each number on the scale containing a detailed descriptions on what physical feelings corresponds to the levels of effort. For Map runner it was somewhat unclear what it wanted to capture. The header pointed out it was capturing "training feeling" but the scale ranged from exhausting to easy, indicating it focused on capturing exertion.

In nearly all analysed apps, once a subjective metric is entered and the training session is saved, there is no further integration of the subjective data with other types of data. Neither is there any guidance to make sense of the captured subjective metric. The subjective metric can only be accessed by clicking on a workout in the list of activities and seeing its details.

Strava, Training Peaks and Vortza are the exceptions to this and do provide interpretation. Strava compares the Relative Effort to previous workouts and showing how this measure aligns with the average Relative Effort (see Figure 2). It is important to note that in Strava, Relative Effort is based on either RPE or cardiovascular data. This depends on what kind of data is available and what the preference is of the athlete, meaning that Relative Effort could be either an objective or a subjective metric. As states in Figure 2, the Relative Effort can help determine whether training is on track or if rest is needed, providing guidance to insights for the athlete. Although, this feature is mainly behind a paywall, restricting access for some users.

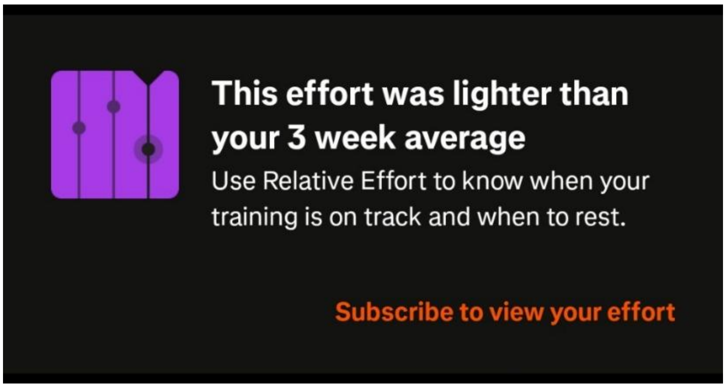


Figure 2
Strava's Insight on RPE

Training peaks provides a raster between the two subjective scales it captures (RPE and feeling) to give advice for further training [4]. For example, a very high RPE and an unpleasant feeling indicates the athlete should reduce training load.

Among the analysed apps, only Vortza and Strava integrate the subjective measures with objective training data. Vortza uses the RPE data to make changes in training plan and automatically updates the next planned training. When RPE is reported very low, next training will be more intense and a high RPE will result in next training being less intense than initially planned. Strava integrates the Relative effort with training data to calculate three key scores: fitness, fatigue and form[5]. (However, subjectivity is only factored into these scores when the RPE is used for Relative Effort.)

- Fitness score: Reflects the overall volume and consistency of exercise. If a person maintains a steady exercise routine, their fitness score remains stable. If they stop exercising, the score gradually drops.
- Fatigue score: Increases immediately after a workout, representing post-exercise exertion. It decreases rapidly with rest.
- Form Score: Calculated as the difference between Fitness and Fatigue. When fitness is high and fatigue is low (after adequate rest), the form score peaks, indicating that an athlete is in optimal condition.

Strava bases these calculations on the impulse response model of Calvert et al (1976). Relative Effort plays a role in calculations for the fitness and fatigue scores, since a workout that is perceived as difficult contributes to both a higher fitness and a higher fatigue. Interestingly, the naming of these scores can be misleading, as they refer to feelings experienced through bodily sensations. Terms such as Fitness, Fatigue, and Form may suggest that these scores are purely subjective, which can create confusion when an athlete's personal experience does not align with the calculated values. Bentvelzen et al. (2022) explain that such derived metrics can be difficult to interpret, often leaving it unclear what the scores actually represent. They argue that this lack of clarity can lead users to question the accuracy of the data and lose trust in the system. Given that these scores can be based entirely on objective measures (when cardiovascular data is used for relative effort), this confusion is a real possibility. Training peaks also includes a fitness, fatigue and form score but does not use any subjective data for calculating these.

From this analysis we learned to what extent subjective metrics are integrated into self-tracking applications for running. Apart from Strava, Vortza and Training peaks, none of the other apps make an effort to help users understand their subjective data or show how it connects to other

[4] https://www.trainingpeaks.com/learn/articles/what-are-rpe-and-subjective-feedback/
[5] https://stories.strava.com/articles/how-to-use-stravas-fitness-and-freshness-tool

performance metrics. On the website of Garmin Connect it is explained that "Self-evaluations are for personal use and do not feed into training or performance metrics, and are intended only to provide a way to subjectively track personal progress when completing similar workouts over time." [6]

Research by Palsa and Mertala (2023) confirms that many running apps fail to connect objective and subjective metric, an observation that aligns with our own findings. Additionally, the analysed apps only tracked RPE or a simple feeling scale, which creates a clear opportunity for us to introduce new metrics that could provide new meaningful insights. We identify Strava and Vortza as having the most sophisticated implementations of subjective metrics. However, Vortza only uses it for managing training load and Strava incorporates subjective data only when RPE is factored into the calculation for Relative Effort. As a result, the experiential aspect of workouts can still be overlooked.

In conclusion, there are clear opportunities to integrate subjective metrics that are meaningfully linked to objective data and designed for interaction at different times and contexts, beyond just post-run input.

2.2 Sports Science Literature on Subjective Metrics

While the previous section demonstrated that apps often incorporate some level of subjectivity, such as RPE or the feeling scale, it remains unclear what these metrics truly measure and how they are relevant to a runner. To this end, we explore the literature on RPE, examining its meaning and interpretation, while also covering other relevant metrics and the perspective on metrics.

2.2.1 Rate of Perceived Exertion (RPE)

For the past 50 years, researchers have been studying performance and perceived exertion in physical activity (Williams, 2017). The way someone experiences exertion is personal and can help estimate how hard the body is working. Understanding work intensity is important because pushing beyond one's limits can lead to injuries or health issues, especially when job demands exceed a worker's physical abilities. The most widely used scale is The Borg Rating of Perceived Exertion (RPE) and is a tool used to measure how hard their body is working based on physical signs like a faster heartbeat, heavier breathing, sweating, and muscle fatigue during physical activity (Williams, 2017). The Borg scale aimed to be highly correlated with the heart rate of the individual using it. It ranges from 6 (no-exertion) to 20 (maximal exertion) which correlates to a heart rate of 60 beats per minute (BPM) in rest and 200 BPM for extreme exertion. Naturally, this correlation with BPM does not hold for every individual. More on this in the next section.

Ritchie (2012) explains that RPE can be applied in two modes: (1) In estimation mode, the individual provides RPE ratings during activities to monitor exercise tolerance. It is often measured along with objective measurements such as heart rate or ECG. (2) In production mode RPE is used to controls the intensity of exercise, with ratings corresponding to low, moderate, and high intensity intervals. The athlete can adjust their pace during exercise to ensure that their effort matches the prescribed intensity. Scheid and O'Donnell (2019) explain that in running, this production mode is often used to estimate the maximum heart rate. When estimating maximum heart rate, athletes gradually increase their effort until they reach their (near) limit, at which point their corresponding heart rate is recorded. Knowing the maximum heart rate, helps to improve the use of the Borg scale, which typically ranges from 60 to 200 BPM. Since individual heart rate ranges vary, it is important to know the personal maximum heart rate such that the scale can be adapted slightly. RPE is often compared to heart rate, blood lactate concentration and maximal oxygen uptake as these has some

[6] https://support.garmin.com/en-US/?faq=8nISJXqSZVAI3Td4IWRqsA

statistical correlation (Zinoubi et al., 2017). However, the correlation is stronger in production mode than in estimation mode (Ritchie, 2012).

Like most subjective scales, there are considerable inter-individual variables that must be considered when using them (Ritchie, 2012). Individual ratings can be influenced by age, environmental conditions, mood states and other psychological factors. Therefore, comparing scores between individuals is not particularly meaningful, as factors can affect each person differently. Is it equally important to have proper training for users who use the scales. For example, it should be clear that fatigue should apply to the whole body instead of just a local body part.

When exercising for a long time at a steady pace, RPE gradually increases the longer you exercise, largely due to accumulated fatigue (Pires et al., 2011). Researchers have observed that this increase in RPE tends to follow a predictable pattern: it rises steadily as you approach the end of the workout, regardless the length of the workout (Eston et al., 2012). Whether you run for 30 minutes or 90 minutes, RPE increases consistently in proportion to how much of the workout is completed, not the absolute time that has elapsed (Swart et al., 2009). Noakes (2008), suggest that individuals anticipate the total duration of exercise and plan the increased effort in advance. From the moment exercise begins, they seem to consider how long the session will last and adjusts RPE, and thus speed (pace), based on the anticipated time or distance. This adjustment is influenced by previous experience, current energy levels, and environmental conditions such as heat or altitude (Swart et al., 2009). Consequently, the rate at which RPE increases can be used to predict how much longer an athlete can maintain a given pace.

Eston (2012) explains the Hazard score and the Estimated Time Limit (ETL) scale, which are two more advanced scales that build upon RPE for better pacing and endurance regulation. The Hazard score combines an athlete's instantaneous RPE with the percentage of the event remaining, offering a dynamic measure of pacing strategy. This allows athletes and coaches to anticipate when performance may decline due to accumulating fatigue and make informed adjustments to effort. For example, if a runner reports an RPE of 8 with 60% of the race still to go, the Hazard score would indicate a potential risk of fatigue setting in too early, suggesting the need to reduce intensity to conserve energy for the latter stages. Similarly, the ETL scale is a valuable tool that adds a predictive layer to subjective assessments of exertion. While RPE indicates how hard an activity feels at a given moment, the ETL scale translates that perception into an estimate of how much longer an individual can sustain that exertion. For example, if a runner rates their exertion as a 7 out of 10, the ETL scale might indicate that they can maintain that pace for about 20 more minutes. This insight allows athletes to more effectively adjust their pace, adjust intensity mid-workout, and avoid premature fatigue.

2.2.2 Alternative Subjective Metrics

Research by Grant et al. (1999) compared the Borg Rating of Perceived Exertion scale to other popular subjective scales, including the Visual Analog Scale (VAS) and the Likert scale. The study found that the VAS is widely regarded as one of the most effective tools for measuring subjective experiences such as pain, breathlessness, general fatigue and perceived effort. Its continuous nature, where respondents mark a point on a line representing a range from "no pain" to "worst pain imaginable" or from "no exertion" to "maximal exertion," provides a high level of sensitivity and granularity. This makes it particularly effective in capturing subtle variations in perception that may be overlooked by other scales (Grant et al., 1999). While the Borg RPE scale, which uses numerical values (6 to 20), is structured and commonly used in physical activity settings, it is less flexible and precise compared to the VAS. The Likert scale, on the other hand, relies on discrete statements to assess attitudes or feelings, which can be less suitable for accurately measuring the immediate intensity of exertion or pain. The VAS's ability to provide continuous data with minimal bias makes it an ideal tool in clinical settings, pain management, and sports science (Grant et al., 1999). Numerous studies have shown that the VAS offers superior accuracy in measuring subjective intensity levels, often outperforming other scales in terms of reliability and user-friendliness.

For other alternative subjective metrics, we take inspiration from research by Den Hartigh et al. (2022) that focuses on resilience in athletes. While their research addresses a different fundamental problem, it remains relevant to our work because it integrates subjective experiences (called 'mental metrics') with objective data to provide meaningful feedback to the athlete. Among RPE, their research also uses metrics for mood, motivation and self-efficacy, as a key indicator of resilience. Correspondingly, Saw et al. (2015) explore how both subjective and objective measures can be used to monitor athletes' wellbeing.

One widely used self-report tool is the Profile of Mood States (POMS), which assesses an athlete's mood by measuring six emotional states: excitement, depression, anger, fatigue, confusion, and vigour (Ralph, 1993). Athletes rate how they have felt during a given period using a five-point Likert scale. The total mood disturbance (TMD) score is calculated by adding the negative mood scores and subtracting the vigour score. A higher TMD indicates greater emotional distress, possibly signalling excessive training stress or fatigue. POMS is often used to track mood swings over time, helping coaches identify early signs of overtraining.

Another comprehensive tool is the Recovery-Stress Questionnaire for Athletes (RESTQ-S), which measures the balance between stress and recovery in both training and daily life (Kellmann & Kallus, 2001, pp. 76–136). This tool includes various subscales that assess general stress, emotional fatigue, physical recovery, sleep quality, and overall well-being. Athletes rate the frequency of specific emotions or symptoms experienced over the past few days. RESTQ-S helps distinguish stress resulting from training versus external personal factors. A high stress score paired with a low recovery score may indicate a heightened risk of overtraining, allowing teams to make informed decisions about adjusting training loads.

The Daily Analysis of Life Demands for Athletes (DALDA) is a quicker, more straightforward self-reporting tool used to monitor athletes' perceived stress and fatigue on a daily basis (Rushall, 1990). It is divided into two sections: one that evaluates general life stressors and another that focuses on sport-specific stress. Athletes categorize each item as 'better than normal', 'normal', or 'worse than normal'. A high number of 'worse than normal' responses can suggest that an athlete is under excessive stress or fatigue. Due to its simplicity and ease of use, DALDA is frequently administered daily to track short-term wellbeing and facilitate immediate training adjustments.

POMS and RESTQ-S are effective tools for tracking overtraining, fatigue, and stress, but their complexity and time demands can discourage athletes from using them, especially if the benefits are not immediately clear. Research indicates that people are less likely to complete lengthy surveys when they do not see direct personal benefits (Rolstad et al., 2011). This discouragement can be explained by respondent fatigue, a phenomenon where survey participants lose motivation and attention over time, leading to lower-quality responses (Ben-Nun, 2024). As surveys become longer or more complex, respondents may skip questions, provide superficial answers, or resort to repetitive response patterns, ultimately compromising the accuracy of the collected data. In contrast to POMS and RESTQ-S, DALDA offers a simple approach, making it easier to implement in quick interactions. Although it lacks the depth of POMS and RESTQ-S, its simplicity encourages consistent use.

Den Hartigh et al. (2022) highlights some lessons learned from their project that could inform this research. They emphasize the role of data science, particularly machine learning, in analysing trends and predicting subjective metrics based on objective data. Discrepancies between predicted and reported values could signal important changes for the athlete. However, applying machine learning in our study may not be feasible, as it requires extensive data and more time than is available. Secondly, they highlight the risk of response bias, which can occur if athletes believe their reported data affects something they value, for example, if they think lower fatigue scores will lead to more intense training plans. When designing our intervention, it's crucial to consider such unintended consequences, as athletes may misreport subjective metrics to achieve a perceived benefit.

2.2.3 Perspectives on Metrics

Doherty and Doherty (2018) explain how self-reporting is a complex method of capturing human experience, which is shaped by the interplay between how we perceive, remember and anticipate events. They explain how experience can be viewed through different lenses which they call 'different selves', each offering a unique perspective. The experiencing self is concerned with the immediate moment, shaped by attention, social contexts and bodily sensations, but it does not retain these experiences for long. In contrast, the remembering self takes a retrospective view on events and constructs narratives, influenced not only by the event itself but also by cognitive biases, norms and values. This self often distorts reality, emphasising peak moments while neglecting the duration of an experience. The future-oriented self, responsible for imagining possibilities and making plans, relies on the reconstructions of the remembered self to shape expectations and anticipate future pleasure. Self-reporting on the same event can produce drastically different results depending on which of the selves produces the response. Asking someone if they are enjoying an experience right now will produce different results than asking if they will enjoy it in the future or if they enjoyed it in retrospect. Doherty and Doherty (2018) describe this discrepancy as inter-self-dissonance, highlighting how different selves interpret the same experience in conflicting ways. They emphasise the importance of considering all the different perspectives in the design of self-report tools.

Postma et al. (2022) explore the design space of sports interaction technology in which a similar perspective is given. Among other topics, they examine the timing of user interactions with technology. They categorise these interactions into three distinct modes: **prospective**, which occurs before the activity and captures expectations and motivation; **conspective**, which takes place during the activity and provides real-time insights into factors such as effort and discomfort; and **retrospective**, which occurs after the activity and relies on memory to evaluate the experience. This view complements the work of Doherty and Doherty (2018) and offers an approach for designing self-reporting within sport technology.

Considering the analysed running apps and their implementation of subjective metrics in the previous section, we conclude that all the analysed applications rely solely on a retrospective interaction. Therefore, only the "remembering self" is reporting on the experience. We see opportunities in designing a subjective metric that is interacted with in a prospective or conspective way, gaining more perspective from the experiencing- and future-orientated self.

Postma et al. (2022) also describe the "Form/Space" of an interaction, which refers to where it takes place. As noted in the section 2.1, all the analysed apps collect subjective metrics immediately after the run, likely in the same location as where the activity took place, such as on a treadmill or running track. This is what Postma et al. (2022) refer to as an in-situ interaction. However, their research also suggests the potential value of ex-situ approaches, where users provide feedback in a different setting. For example, subjective data could be collected a few hours after exercise, during the recovery phase at home.

2.3 Identifying an Aspect of Experience

In this section we will highlight literature that emphasises the importance of subjective metrics in enhancing the experience of the athlete. In combination with the previous chapters, we can make an informed choice about the specific aspect of experience we will focus on in rest of the thesis. Thus, answering the first research question: “What opportunities exist for integrating a subjective metric into self-tracking technology to influence the runner's experience?”

Tholander and Nylander (2015) describe how athletes experience both a measured sense and a lived sense of performance, highlighting the importance of a balance between these two perspectives. The measured performance refers to objective, quantifiable data such as heart rate, GPS tracking and pace provided by tracking technology. This data allows athletes to track their progress, compare results and optimise their training based on objective indicators. On the other hand, the lived sense of performance is rooted in subjective experiences. How athletes feel during training, such as their bodily sensations, perception of effort, fatigue and flow is part of this. The study shows that many athletes seek a balance between these two perspectives, using data as a reference while listening to their bodies to guide their training. Since our research focuses on integrating subjective metrics, we see an opportunity to use them as a tool to help athletes better tune into their bodily sensations during training.

Research by Karahanoğlu et al (2024) on TLM emphasises the critical role of integrating subjective experiences and body cues into sports technology. The study identifies two groups among users of self-tracking technology: those in the guided TLM group, who rely primarily on technology for advice and recommendations, and those in the self-directed TLM group, who have developed a more intuitive understanding of their somatic sensations and can effectively manage their training load based on these cues. This research highlights the importance of fostering a deeper connection between athletes and their own bodily sensation, as blindly following TLM advice without considering these internal signals can lead to overtraining, injury, and a misalignment between the athlete's true physical state and the technology's recommendations. Our intervention can focus on supporting athletes in the guided TLM group, helping them to transition to a self-directed TLM approach by encouraging greater awareness of their internal cues and improving their ability to interpret and act on them during exercise.

This process of tuning into internal bodily signals aligns closely with the concept of interoception, which refers to the brain’s ability to sense internal bodily signals, such as heart rate, breathing, and fatigue (Wallman-Jones et al., 2021). This awareness is essential for regulating physical effort and preventing overexertion. One model that explains how this works is called the interoceptive predictive coding model (Seth et al., 2012). It suggests that the brain uses past experiences to create an internal idea, or "template", of how much effort an activity should take. As the athlete starts exercising, the brain compares what the body is feeling right now to this template and if the feelings do not match the expectation, it creates a "prediction error", which tells the brain that something is off (Seth et al., 2018). When this happens, the athlete has two options. It can either change the body’s actions to match the template or it can change the template to better fit what the body is actually feeling (Seth et al., 2018). For example, if an athlete planned to run 10 km but feels more tired than expected at the halfway point, they might slow their pace or adjust their expectations for how tough the run will be. This shows that how tired we feel is not just based on raw data from the body, but also on what our brain thought would happen (Wallman-Jones et al., 2021).

Desmedt et al. (2023) propose two main directions of information flow between the brain and body: bottom-up and top-down. Bottom-up processing refers to the flow of sensory signals from the body to the brain, such signals contain heartbeat, heavy breathing, or muscle soreness. Exercise enhances these signals, which may train the brain to become more sensitive and accurate in interpreting them. On the other hand, top-down processing involves the brain's regulation of how much attention is directed toward these bodily cues. Some individuals may be distracted and overlook them, while others may be highly attuned and attach specific meaning to these sensations.

Research suggests that physical activity can improve interoception by both amplifying bodily signals and training the brain to attend to them (Wallman-Jones et al., 2021). We see an opportunity for our intervention to enhance interoception by nudging the athlete to become aware of the body during running and thereby situating the top-down information flow. Designing technology that enhance awareness of bodily signals requires the development of systems that engage with athletes in real time during their workout, offering immediate feedback. This approach aligns with the opportunity discussed in section 2.2.3, where we uncovered the potential of designing a conspective interaction. This leads us to the following directions:

Focusing on implementing a **conspective subjective metric** with the goal to enhances the athlete's **interoceptive awareness.**

In addition, we identified another direction to explore. Karahanoğlu et al (2024) highlight that for TLM, athletes often prioritise their bodily signals over advice from self-trackers, such as recovery times, because they perceive these recommendations as unreliable. The study suggests several reasons for this, including inaccuracies in objective data, lack of transparency, and concerns about the validity of TLM calculations. In addition, athletes' trust in the advice is affected by whether the recommendations are in line with their personal feelings. Interestingly, Strava's implementation of the fitness, fatigue and form scores, discussed in section 2.1, is a good example of where these problems possibly occur. As described in that section, the naming of the scores can stimulate the comparison of the scores to the athletes' personal feelings. Hence, a mismatch between the scores and the experienced feelings can lead to a lack of trust. In addition, Strava can calculate the scores solely on heart rate data, without any subjective input being considered. We emphasise that objective data alone does not provide a complete basis for effective TLM guidance. Therefore, we envision that integrating subjective measures could increase confidence and offer a more reliable basis for TLM counselling. Given the design opportunities identified in section 2.3.3, we also see potential in implementing an ex-situ subjective measure.

Focusing on implementing an **ex-situ subjective metric** with the goal of enhancing the athlete's **trust** in TLM advice.

To answer the research question, 'What opportunities exist for integrating a subjective measure into self-tracking technology to influence the runner's experience?', we identify two key opportunities. Firstly, implementing a **conspective** subjective measure to increase athletes' awareness of their body sensations during exercise. This would allow them to better tune into their internal signals and improve training load management e.g. Second, integrating an **ex-situ** subjective metrics to increase confidence in TLM by adjusting advice using athletes' personal feelings. However, due to time constraints, we cannot work them both out and therefore decided to focus on the first opportunity.

3. Designing the Subjective Metrics

In this chapter, we focus on designing and implementing the conspective subjective metric within self-tracking technology for runners. For clarity and consistency, we refer to the overall solution as the intervention. This terminology also aligns with the structure of the next chapter, where we evaluate the designed and developed system against a baseline.

To begin, we make a choice on the structure of the intervention, and the devices involved. A smartwatch will be used to query the subjective metric during the run, enabling real-time (conspective) interaction that supports interoceptive awareness. To extend the intervention beyond the run, we also include post-activity interactions on a smartphone, engaging the remembering self explained in section 2.2.3. This division of tasks between devices is supported by Mortazavi et al. (2015), who highlight that wearables are more effective during physical activity, while smartphones are better suited for data analysis and reflection. This approach is also adopted by the larger fitness watch brands like Polar and Garmin[7] [8].

In the following section, we use the work of Postma et al. (2022), who outline a design space for sports interaction technology. Postma et al. (2022) addresses both the form and the function of the interaction in their design space. Since the function of our intervention is quite already clear, the focus will mostly cover aspect related to the form of the interaction. This framework helps ensure that all relevant design aspects of the intervention are considered. These aspects are noted down as requirements and will guide both the ideation session and detailed design decisions. At the end of section 3, we present the most promising concept as the final intervention.

3.1 The Form of the Interaction

3.1.1 Space

The interaction space refers to where the interaction with the intervention takes place (Postma et al., 2022). This can be either in the environment where the activity takes place (in-situ) or outside of it (ex-situ). In our situation, we envision the interaction to take place in both spaces. The interaction with the smartwatch takes place during the run, making this an in-situ interaction. Postma et al, (2022) suggests that in-situ interactions enhance learning because the scenario is closest to the normal activity (without the intervention). The smartwatch can facilitate real-time engagement by providing runners with subjective metrics and feedback focused on understanding their bodily sensations. The smartphone will be used outside of the run for preparation and reflection (ex-situ). Athletes can review their subjective responses from previous runs, compare them to objective performance measures, and adjust their training accordingly. The smartphone's larger screen and advanced computing power offer an efficient way to analyze data and track long-term trends, making it well-suited for ex-situ use. In a real-world scenario, athletes might review their run immediately after finishing, such as during their walk home. This quick reflection may be somewhat rushed, therefore worsening the reflection. Encouraging a dedicated ex-situ space for deeper reflection could be beneficial. One way to achieve this is by sending a notification an hour after the activity, prompting athletes to review their run with a fresh perspective. Although ex-situ interaction might have a smaller transfer effect, meaning the skills developed may not fully carry over into real-time performance, serving as an addition to regular training is still beneficial (Renshaw et al., 2019).

[7] https://connect.garmin.com/
[8] https://flow.polar.com/

3.1.2 Time

The timing of the interaction is about when the interaction takes place in relation to the activity. In accordance with Postma et al (2022), we can distinguish between three temporal aspects, before (prospective), during (conspective) and after (retrospective) the activity. Before the run, we expect the smartphone to be the primary device users interact with, given its larger screen and greater usability. During this phase, users should receive a thorough onboarding, explaining how the intervention will function. Specifically, clear expectations should be set regarding how the smartwatch will operate during the run, including what prompts will appear, why they are relevant, and how the athlete should respond. The core of the intervention occurs in a conspective way, which mean the athlete will interact with the smartwatch during the activity. At certain times, e.g. at regular intervals or in response to physiological changes such as increased heart rate, the smartwatch can ask runners subjective questions. This real-time engagement can help runners tune into their bodily sensations, increasing awareness. Post-run reflection occurs on the smartphone, leveraging its usability for deeper analysis. Athletes can review their subjective responses alongside objective performance data, such as pace, heart rate, and distance. This retrospective analysis provides a deeper understanding of how bodily sensations correlate with performance measures over time. By integrating these multiple dimensions, the intervention provides a comprehensive approach to incorporating subjective metrics into self-tracking techniques. Stimulating the athlete to think about the activity from all dimensions of time is also deemed as important by Doherty and Doherty (2018)

3.1.3 Nature

The nature of the interaction is about by how the intervention changes the experience (or nature) of the sport. Postma et al. (2022) highlights factors such as exercise modification, behaviour steering, constraining action and gamification, playification, sportification, which will be discussed below.

Exercise modification refers to how technology influences a sports activity, ranging from minor enhancements to complete transformations (Postma et al., 2022). According to Ishii et al. (1999) and Postma et al. (2019), this modification can be categorised into three levels.: **Inform**, where the activity remains unchanged but provides additional data to the athlete; **Augment**, where interactive elements enhance the experience while keeping it similar to the original activity; and **Transform**, where the technology creates an entirely new sports experience. In the context of smartwatch and smartphone use, informing occurs when a smartwatch provides real-time metrics like heart rate, encouraging athletes to focus on bodily sensations, while the smartphone later presents reflections alongside objective data for deeper insight. Augmenting happens when the smartwatch actively engages athletes by prompting subjective metrics during training, guiding them to shift their focus inward. However, the intervention does not aim to completely transform the act of running, as maintaining its core nature remains essential.

Behavior steering refers to how technology influences an athlete's actions, attitudes, and perceptions, ranging from subtle nudges to direct control (Postma et al., 2022). Research by Delden et al. (2014) explains how this influence can be assessed through three key dimensions: **Forcefulness:** how much freedom athletes have in deviating from the system's guidance; **Transparency:** how aware they are of being influenced; and **Validity:** the accuracy and reliability of the information provided. In our intervention, we see value in minimal forcefulness, ensuring that athletes are in control by allowing them to skip questions and adjust the frequency of prompts for example. This avoids annoyance towards the system. Next, we anticipate for the intervention to be highly transparent. It is crucial for athletes to understand why the subjective questions are asked and how participating in the system is valuable to their own wellbeing. Finally, we see no reason to change the validity of the information. Therefore, we will leave it uncompromised, ensuring that all feedback is trustworthy and as accurate as possible. As for behaviour steering approaches mentioned by Postma et al. (2022), we envision a balance between guidance with autonomy, relying on **enticing**, by using rewards to encourage behavioural changes, and **coaxing**, by subtly presenting relevant information to guide decision-making without imposing strict control.

Constraints on Action refers to how technology can influence or limit the sporting activity (Newell & Jordan, 2007). Newell and Jordan identify three types of constraints that shape how an activity unfolds: performer constraints (e.g., body structure), environmental constraints (e.g., temperature or surroundings), and task constraints (e.g., goals or rules). Our intervention will incorporate a task constraint, but without forcing it on the athlete. While the usual running goals might include completing the run, reaching a certain distance, or maintaining a specific pace, the athlete is also given an additional task, which is to fill in the subjective metric and getting aware of their bodily sensations. This shift encourages learning and self-awareness, while ensuring the athlete retains control over the experience.

Finally, **gamification, playification, and sportification** are methods to enhance engagement and tailor experiences through game-like, playful, or sport-like elements (Schell & Schell, 2008). While this is not the core of the intervention we are trying to make, we do think adding elements could stimulate bodily awareness even more by making answering the subjective questions more rewarding.

3.1.4 Feedback

Feedback in sports technology is about how the technology communicates with the athlete to improve the task they are doing (Postma et al., 2022). The effectiveness of feedback depends on factors such as timing, modality, frequency, and content. In our case, the subjective metric itself is a form of feedback, encouraging the athlete to focus inward on their body's sensations. We also see potential to add more ways to provide feedback, like using notifications to guide attention to their breathing or showing the athlete's heart rate to them.

The timing of the feedback has already discussed in the sections regarding to the space and time of the interaction and will therefore not be mentioned again here.

The modality of the feedback is about which of the human senses is used as a communication channels (Postma et al., 2022). The smartwatch is equipped with a touchscreen, speakers, and a vibrating module, which allows us to provide feedback visually, audibly, and through haptics. Since the athlete needs to fill in the subjective metric, it makes the most sense to present the feedback visually on the touchscreen, where they can also input and submit their responses. Alternatively, the smartwatch's speakers can read the subjective metric aloud and answer could be collected through speech recognition. Although, using speech recognition for input may be less effective in noisy environments or when the athlete is out of breath. The haptic module is not ideal for directly querying the subjective metric but can be used for giving other types of feedback. E.g. simulating the heart rate or provide cues to help the athlete focus inward. A more effective use of haptics would be to draw the athlete's attention to the watch, allowing the interaction to continue visually on the touchscreen. Combining multiple modalities can also enhance communication, reinforcing the feedback in a more engaging way, described as multimodal feedback. On the smartphone, the feedback will naturally be given visually on the touchscreen as well.

There are several strategies to consider when determining the **frequency of feedback** (Postma et al., 2022). One option is to prompt the metric at regular intervals, such as every kilometre or after a set amount of time. Potentially, the feedback can decrease over time when the athlete does not require it anymore. Another option is to let the athlete request the feedback whenever they feel the need to reflect. A third option is to trigger the feedback automatically when the athlete reaches specific physical thresholds, like an increased heart rate, ensuring the feedback is most helpful during crucial moments. For the interaction with the smartphone, we see summary and average feedback as most prominent. Summary feedback provides an overview of the athlete's performance taking in multiple datapoints, while average feedback focuses on general trends rather than specific moments, making it well-suited for post-run reflection and learning.

The content of feedback has a few important subcategories to consider, and notably, Postma et al. (2022) focus heavily on motor learning and whether feedback should label actions as either 'good' or 'bad' based on how they support skill acquisition. Our intervention takes a different approach, since we are not targeting motor learning, this kind of binary judgment does not apply. Instead, we aim to support learning by helping participants explore and refine their bodily awareness. Because of this, we do not apply every aspect of traditional feedback theory. However, we can make a distinguishment between feedback being either knowledge of results or knowledge of performance. In our case, the feedback is more aligned with knowledge of results, for example, giving insights into how accurate someone's perception of their bodily signals was. Knowledge of performance, which looks at what behaviours led to a certain outcome, is not useful here because we cannot observe the internal processes going on in someone's mind during the run. As for the **precision of feedbac**k, it should be specific and clear during the run, but it can be more general in the reflection afterward. Finally, our feedback will focus on what went right. Research suggests that this positive framing supports self-reflection and helps prevent overthinking or rumination (Niess et al., 2020).

3.1.5 Integration

The integration of the intervention is about how it fits into an athlete's training routine, taking into account elements such as frequency and context (Postma et al., 2022). Defining the exact pattern of use is not the main focus, but a valuable consideration is how mental training can be incorporated. Weinberg and Gould (2023) emphasize that mental training can be effective as an addition to physical training. For example, a retrospective interaction with the smartphone can make athletes mentally go back to specific moments during the run. This technique, often called internal imagery, helps athletes reconnect with the physical and emotional sensations they experienced (Weinberg & Gould, 2023). By engaging more deeply with these memories, athletes can improve their bodily awareness and add more personal meaning to their post-run reflections, ultimately making the experience more impactful and insightful.

3.2 The Function of the Interaction

Although the function of the interaction is rather clear, it is still valuable to consider some aspect and see how the intervention can be categorised using different lenses.

3.2.1 Application Scope

The application scope defines the goal of the interaction (Postma et al., 2022). Our intervention primarily falls in the physical education scope by promoting interoceptive awareness, helping athletes to make better decision and prevent them from overtraining and injury. In addition, the intervention can increase engagement in sport as a leisure activity. By e.g. incorporating gamification elements or interactive feedback, it can enhance the hedonic aspects of exercise, making the experience more enjoyable and intrinsically motivating.

3.2.2 Nature of Sport and its Training

The nature of sport and its training plays a crucial role in shaping how athletes improve and stay motivated. Postma et al. (2022) identifies several categories from which a few are interesting to consider.

First al all, we see a facilitative function of the feedback as most prominent, meaning that the desired effect of the intervention can still be achieved without it, for example through years of running experience as point out by Wallman-Jones et al. (2021). However, the feedback should still produce a positive effect while avoiding hindering other positive effects. As some runners may not want any distractions during running as stated in research by Tholander and Nylander (2015), it is important to consider how this intervention can work for them and whether it could be effective at all for individuals who prefer a more uninterrupted experience.

The temporal structural aspects highlight the importance of considering different phases of training. In running, as a race comes closer, training typically changes and becomes more focused. Although the intervention will not be designed for a specific training stage, in real-world scenario's it is important to consider how such an intervention fits into the training plan of athletes.

Freedom in functionality in sports interaction technology emphasises the difference between closed and open design. A closed design is one that is limited to a specific scenario and offers little or no customisation for the user. In contrast, an open design gives users more flexibility, allowing them to adapt the technology to their preferences. For our testing purposes, we are likely to opt for a more closed design, ensuring that the intervention produces consistent results for each participant. However, we recognise the value of incorporating more open-ended design elements, as this would make the technology accessible to a wider range of users. Possible ways to make it more open include allowing users to decide how often subjective metrics and feedback are provided and offering the option to adjust the questions to better suit their individual needs.

3.2.3 Pedagogy, Learning and Didactics

Pedagogy, Learning, and Didactics touches upon key educational principles and covers subjects such as models-based practices, modelling and learning phases (Postma et al., 2022). Model-based practices are a pedagogical approach that aligns learning outcomes with students' needs and instructional styles (Casey & Kirk, 2020). These practices provide a structured "blueprint" for organizing content, task structures, and the sequencing of learning activities in physical education. In our intervention, we see possibilities to incorporate principles from Teaching Games for Understanding (TGfU), a model that emphasizes game-based learning and decision-making (Hortigüela Alcalá and Hernando Garijo ,2017). The TGfU model points out that letting the students decided which datapoints are most important can enhance engagement and learning. Additionally, critical questioning and reflection are also essential elements of the TGfU approach (Hortigüela Alcalá and Hernando Garijo, 2017). The intervention could use critical questions such as, "Why did I report this RPE" or “What did you feel that made you slow down?". This reflective process fosters deeper learning related to their awareness of bodily sensations by helping participants evaluate their decisions.

Modelling, while beneficial in motor learning, is not applicable to our intervention since an individual's perception of their own body is inherently personal and cannot be externally replicated. However, **learning phases** should be taken into account. The learner's stage of development plays a crucial role in how information is processed, and it is essential to present content in a way that aligns with their cognitive capacity (Edwards, 2010). Providing too much information at once can overwhelm participants, making it difficult for them to absorb and apply key insights. This is also something to consider while in the testing phase, as some participants may already be highly attuned to their bodies, (without realising it) while others may require more guidance to develop this awareness.

3.3 Requirements

Based on the analysis in Sections 3.1 and 3.2, we identified a range of requirements that informed the ideation process, and the final intervention design presented in the following sections. These requirements are categorized using the MoSCoW method, which indicates their priority level (Ahmad et al., 2017). Notes that not all the non-essential requirements are (completely) addressed in the intervention. Table 2 presents an overview of the categorised requirements including the section from which it was derived.

Table 2

Intervention requirements

Priority	Requirement	Section
Must	A smartwatch is used for the in-situ interaction	3.1.1
Must	A smartphone is used for the ex-situ interaction	3.1.1
Won’t	Smartphone gives a notification one hour after the run to ensure ex-situ interaction	3.1.1
Should	All the temporal aspects, (prospective, conspective, retrospective), are addressed in the intervention	3.1.2
Must	The intervention modifies the run in an informing and augmenting way	3.1.3
Must	The intervention should steer behaviour in a forceless way, be transparent about its intent and display untampered information	3.1.3
Should	The intervention should steer behaviour by using an enticing or coaxing approach	3.1.3
Must	The intervention should have a task constrain (using the subjective metric)	3.1.3
Could	The intervention incorporates gamification, playification and sportification elements	3.1.3
Must	The intervention should communicate feedback using multiple modalities	3.1.4
Should	The feedback includes Knowledge of Result	3.1.4

Should	The feedback is precise and clear during the run and can be more general in the reflection	3.1.4
Could	The intervention includes internal imagery elements	3.1.5
Must	The feedback has a facilitative nature	3.2.1
Won't	The intervention has an open design	3.2.2
Should	The intervention integrates aspect of the TGfU model	3.2.3
Should	The intervention prevents information overload during the run	3.2.3

3.4 Ideation

3.4.1 Brainstorm

To generate ideas systematically, we employed the brainwriting technique described by VanGundy (1984). This involved setting a fixed time limit during which we wrote down as many ideas as possible without evaluating them. Input for this session included literature on subjective metrics (see section 2.3) and the exploration of the sports technology design space (see section 3.1). This approach ensured our ideas are both varied and anchored in existing research.

After the session, we selected the three most promising ideas, which was partly based on perceived novelty and feasibility, as well as the extent to which each idea was grounded in the literature and aligned with the design opportunities identified earlier. Refined versions of these three ideas are presented below in Table 3.

Table 3

Intervention ideas

Title	***1. Change in Exertion with DALDA Metric***	***2. Understanding signs of exertion***	***3. On-demand heart rate training***
Main interaction	Periodic question during run: "Lighter", "Similar", or "Heavier" effort in relation to previous reflection point.	Periodic question "What feels like the biggest sign of effort right now?"	Runner requests HR; prompted first with: "What do you think your heart rate is?". Three options shown

Focus of Awareness	Change in exertion intensity	Change in exertion type (e.g. breathing, fatigue or pain)	Momentary exertion
Feedback Mechanism	Immediate display of current heart rate after response	No direct feedback during run; focus is on selection of sensation	Immediate display of current heart rate
Post-Run reflection	Reflect on subjective responses and corresponding heart rate data	Review evolution of chosen sensations.	None
Design rationale	Simple interface minimizes input burden during movement; inspired by DALDA and bodily awareness literature	Encourages natural self-focus; helps identify early warning signs of overexertion or injury	Focuses on a minimal impact on the running experience

1. **Change in Exertion with DALDA Metric**

Research on the Rate of Perceived Exertion (RPE) scale (Borg, 1982; see section 2.3.1) shows a strong correlation between perceived effort and physiological signals such as heart rate. Understanding this relationship can help runners better manage intensity and prevent overtraining. However, using the standard 6–20 RPE scale during a run is impractical, particularly due to its cognitive load and the fine-grained input required on a small wearable interface.

To address this, we propose a simplified, point-to-point reflection method inspired by the structure of DALDA (Rushall, 1990) but adapted to focus on change in perceived effort rather than absolute intensity. At regular intervals, runners are asked whether their effort feels “lighter,” “similar,” or “heavier” compared to the previous checkpoint. This relative approach reduces interaction complexity while still capturing meaningful subjective data. The response is immediately followed by heart rate feedback, enabling the runner to reflect on how their internal sensation aligns with physiological change, supporting learning and body awareness in real-time.

2. **Understanding signs of exertion**

This concept focuses on injury prevention by helping runners become more aware of which bodily signals are most prominent during exertion. At periodic intervals, runners are asked: “What is your strongest sign of effort right now?”, with options like breathing, muscle fatigue, or joint pain. This directs attention to specific bodily sensations and helps identify recurring discomfort that may signal early stages of overuse or stress-related injury. Unlike traditional exertion tracking tools, this concept does not aim to quantify effort, but instead to support somatic awareness. By surfacing what the runner notices most in their body, it offers a low-barrier way to monitor risk factors and encourage self-regulation during training (Karahanoğlu et al., 2024; section 2.4).

3. On-demand heart rate training

This idea explores reducing the frequency of interaction by allowing runners to request heart rate feedback only when they want it. Before revealing the data, the system asks the runner to estimate how they feel, encouraging a moment of inward reflection. One key design consideration here is the frequency of interaction. As discussed in section 3.2.3, some runners prefer uninterrupted training sessions and are easily disengaged by overly intrusive self-tracking technology. This on-demand model serves such preferences, offering reflective moments without enforcing regular prompts. It's lightweight, user-driven, and supports bodily awareness without undermining the flow of the workout.

After reviewing the three concepts, we selected the Change in Exertion with DALDA Metric because we envision it has the most potential to directly support the development of interoceptive awareness while remaining simple and practical for use during running.

3.4.2 Bodystorm

The selected idea requires the runner to interact with the watch at regular intervals during a run. Finding the right timing between prompts is crucial: if messages appear too frequently, they risk interrupting the runner's flow, while if they're too infrequent, the impact of the prompt may be lost due to fading memory. To make an informed choice about this timing, we conducted a bodystorming session while going on an 8-kilometre run, focused on testing different frequencies for querying the subjective metric. Additionally, the bodystorming exercise helps refine our idea, as physically experiencing a concept often uncovers insights that are difficult to achieve through traditional brainstorming alone (Márquez Segura et al., 2016).

Based on this bodystorming session, we chose a 1-kilometer interval as the most balanced option. Shorter intervals tended to feel cause annoyance with the intervention, while longer intervals made it harder to recall and compare with the previous reflection point. Interestingly, we found that answering the question could take quite some time. Focusing longer on physical sensations resulted in more aspects and details about them. For example, subtle feelings like tension, changes in breathing, or discomfort became clearer. Because of this, we think it is useful to give users of the intervention a heads-up before asking them to respond to the subjective metric, giving them time to tune into their bodies. Concluding, the intervention design should include prompts at 1-kilometer intervals, with a brief heads-up beforehand to allow runners to attune to their bodily sensations.

3.5 Final Intervention

In this section, we present the final design of the intervention. Design decisions were informed by the requirements described in section 3.3 and insights gained from a bodystorm session. The section is organised into two parts, each describing one stage of the intervention: during the run (on the smartwatch), and after the run (again on the smartphone).

3.5.1 Smartwatch Interface during run

For most of the run, the user is presented with a simple overview of their activity, including the elapsed distance and duration, as illustrated in Figure 3. We have intentionally left out common metrics such as pace and heart rate. This decision will be explained later in this section.

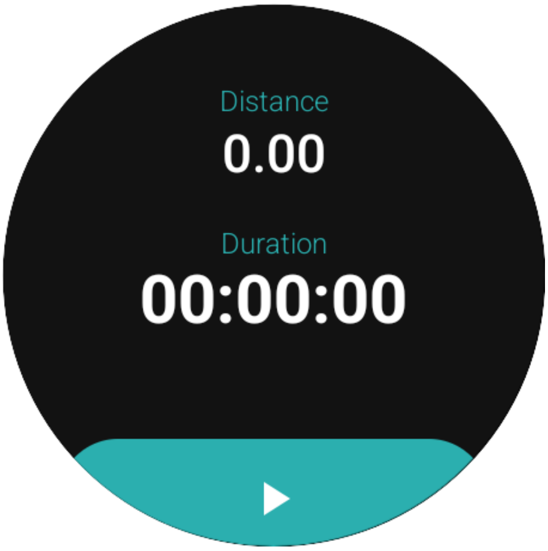


Figure 3
Watch Interface – Main screen

To enhance interoceptive awareness during the run, the smartwatch makes use of nudges to look inwards, questions related to exertion and provides heart rate data. For starters, the watch vibrates after each completed kilometre to draw the runner's attention. At that moment, the watch displays a brief textual message encouraging the runner to look inward and become aware of the sensations within their body, as shown in Figure 4. This prompt is referred to as a Reflection Point. After an additional 100 meters of running, the watch asks the runner whether their body feels lighter, similar, or heavier compared to the previous Reflection Point (Figure 5). Upon receiving a response, the watch then displays the current heart rate alongside the heart rate recorded at the previous Reflection Point, allowing the runner to compare their subjective feelings with the objective data (Figure 7). With this, we aim to engage the "experiencing self", explained by Doherty & Doherty (2018) and help the runner learn which feelings correspond to what heart rate, by connecting subjective and objective measures of exertion.

For this reason, we intentionally left out the average pace from the main screen, as this prevents the runner from simply associating a specific pace with a specific heart rate, which would reduce the need to turn their attention inward and truly feel their body. Similarly, heart rate data has been excluded, as constantly seeing it would make the question irrelevant. In Figure 7, the app shows the current heart rate of 174 and the heart rate of the previous Reflection point, which is 162. The corresponding message says, "You exertion is indeed Heavier." This message is shown because in this example, the user entered "Heavier" in Figure 5, and because the difference in heart rate is more than 5 bpm, which the app interprets as a significant threshold to agree with the users on that the bpm is higher. This threshold of 5 bpm is chosen because we wanted it to be slightly higher than the normal heart rate variability, thus preventing the increase or decrease in heart rate to be a random change (Rajendra Acharya et al., 2006).

The type of subjective metric (lighter, similar, heavier) is inspired by the DALDA self-reporting tool described in section 2.3.2. We selected this approach because the touchscreen is the most practical input method for the runner. Given the smaller screen size and the need for interaction with the device while in motion, it was logical to choose a metric with only a few input options. Offering more response options (e.g., using the RPE-CR10 scale) would require greater precision, making the interaction difficult during running and increasing the risk of accidental inputs. We considered colouring the lighter, similar and heavier buttons for clarity reasons, but decided not to do this as the colouring could introduce a bias in the runner's perception or influence their response. Using colours might subconsciously associate certain options with something positive or negative, which could compromise the objectivity of the self-assessment. For the first Reflection Point, there is no prior data for comparison. In this case, only the current heart rate is shown, and the runner is prompted to consciously connect their physical sensations to this heart rate reading, as shown in Figure 6.

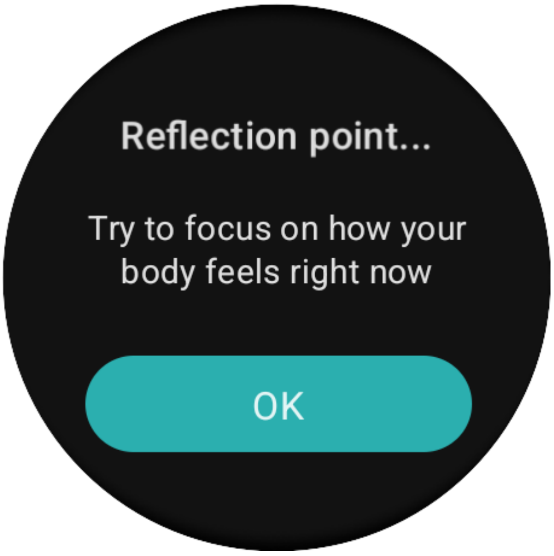


Figure 4
Watch Interface – Reflection point

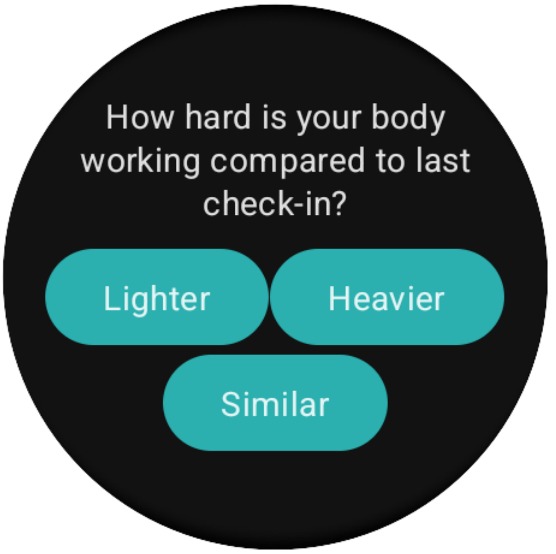


Figure 5
Watch Interface – Subjective Metric

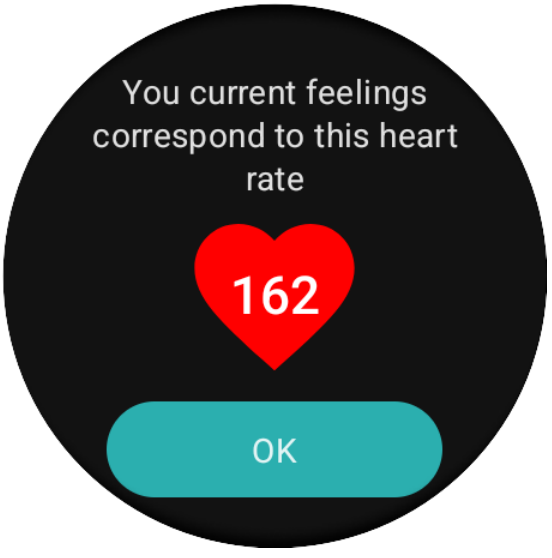


Figure 6
Watch Interface – First feedback

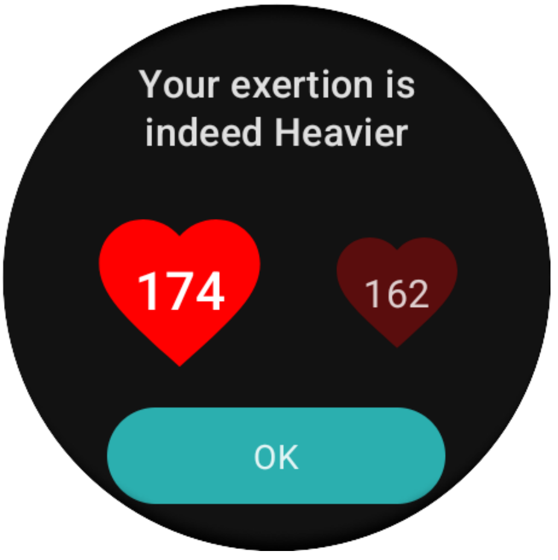


Figure 7
Watch Interface – Reflection feedback

3.5.2 Smartphone Interface after run

For the post run reflection, we focused on implementing the standard elements included in runner apps, taking inspiration from the analysed apps in section 2.1, and complemented this with the subjective metric data captured during the run. In Figure 8, a summary component of the reflection is show, containing elapsed distance, duration, average pace, speed and average heart rate.

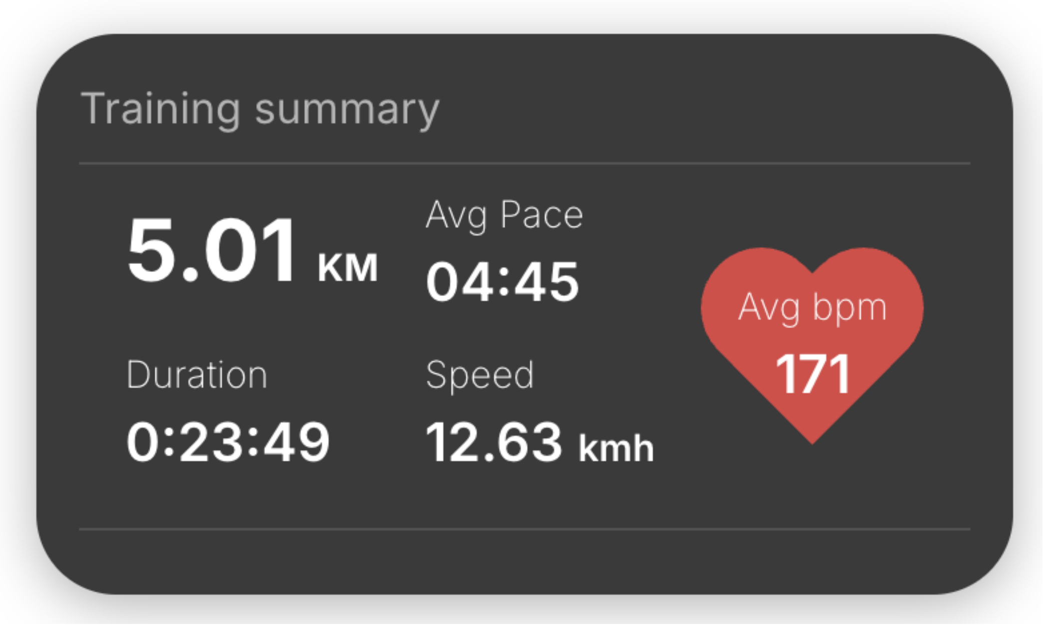


Figure 8
Smartphone Interface - Training Summary

The second element of the reflection is a map component that displays the route the runner has taken (see Figure 9). This map includes markers at each kilometre, which can be clicked to reveal the subjective data recorded on that point, which is shown alongside the corresponding heart rate (Figure 10). As before, the first reflection point only presents the heart rate (Figure 11), while subsequent points also include the subjective metric. This map component gives the runner an opportunity to think back on about the moments in their run in which they were actively sensing their bodies. This visualization helps the runner engage their remembering self, creating a narrative of the run and highlighting moments of heightened bodily awareness (Doherty & Doherty, 2018).

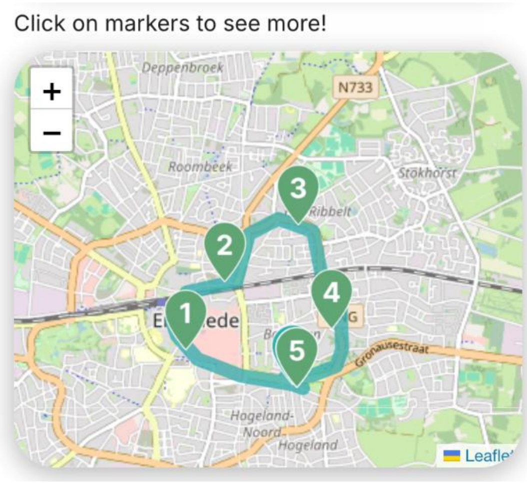


Figure 9
Smartphone Interface - Map

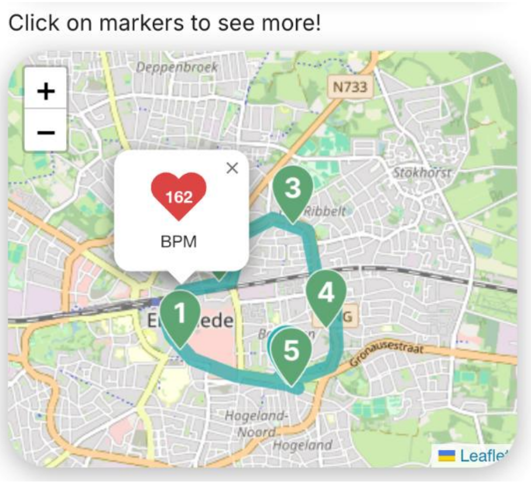


Figure 10
Smartphone Interface - Map First Reflection

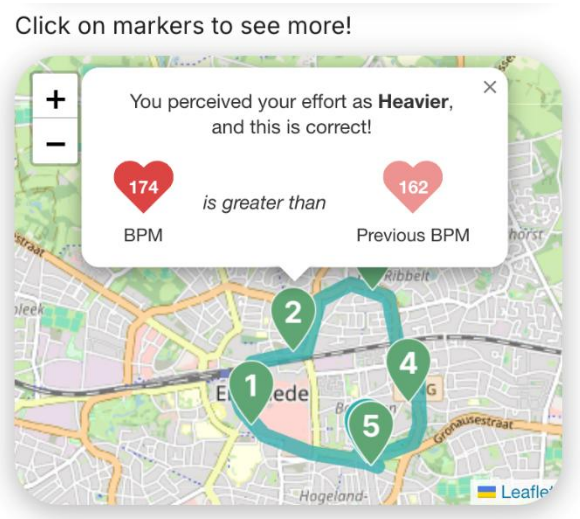


Figure 11
Smartphone Interface - Map Other reflections

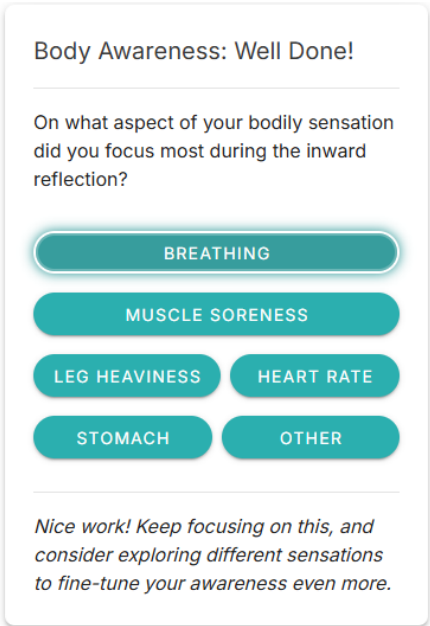


Figure 12
Smartphone Interface – Body Awareness

Figure 12 shows the final part of the post-run reflection. Based on how the participant answered the subjective prompts during the run, they receive tailored messages related to their bodily awareness. These messages are informed by how often their responses matched the change in their heart rate data. A response is considered aligned when the participant's reported change in effort corresponds with a measurable change in heart rate. When this alignment occurs more frequently, it suggests a stronger connection between the participant’s internal sensations and their physiological signals. The messages displayed at the top and bottom of Figure 12 are shaped by this level of alignment and are shown below:

- **Low Alignment**
  Title: “Body Awareness: Room for improvement.”
  Advice: “For your next run, try focusing on a different aspect of your body. Small shifts can already help improve awareness.”
- **Medium Alignment**
  Title: “Body Awareness: Getting there.”
  Advice: “You're on the right track. Try incorporating other aspects like breathing or muscle tension to enhance your awareness.”
- **High Alignment**
  Title: “Body Awareness: Well done!”
  Advice: “Nice work! Keep focusing on this and consider exploring different sensation to fine-tune your awareness even more”

This final component is inspired by the TGFU framework described in section 3.2.3 and the mental training principles in section 3.1.5. In this framework, there is an emphasis on asking critical questions after the workout session, using the captured data as a reference. We applied this approach by asking participants which elements of exertion they reflected on during the designated reflection points. Additionally, they were provided with actionable insights for their next training session to further enhance their self-awareness.

4. Evaluation

In this chapter, we focus on answering the question: "What is the effect of the designed subjective metric on the runner's experience?" To address this, the chapter is structured into three main sections. First, the study design outlines how the research was set up, including the experimental conditions, participant selection, and measurement tools. Second, the data analysis section details the methods used to examine both quantitative and qualitative data, including statistical tests and coding approaches. Finally, the results section presents the key findings from the evaluation, highlighting how the intervention influenced interoceptive awareness, reflection, and user experience. As a preparatory measure, this study has received approval from the Computer & Information Sciences (CIS) Ethics Committee at the University of Twente (see Appendix A).

4.1 Study Design

The study is designed to capture the impact of the intervention on the athletes' interoceptive awareness, type of post-run reflection and overall usability/experience. The evaluation of the intervention will be done using a mixed method approach which follows an A/B test format: Group A serves as the control group (baseline), while group B receives the full intervention. The effect on the athletes' interoceptive awareness will be tested using a quantitative approach using a self-administered questionnaire. Corresponding H_0 and H_1 hypotheses:

Null Hypothesis (H_0):
There IS NOT a significant difference in interoceptive awareness scores between the group that received the intervention (Group B) and the group that did not receive the intervention (Group A)

Alternative Hypothesis (H_1):
There IS a significant difference in interoceptive awareness scores between the group that received the intervention (Group B) and the group that did not receive the intervention (Group A)

We utilised the Multidimensional Assessment of Interoceptive Awareness, version 2 questionnaire (MAIA-2), developed by Mehling et al. (2018). This questionnaire consists of eight distinct subscales, which can be applied selectively depending on the purpose of one's study. Given the expected limited sample size, we focus on the four subscales, Noticing, Non-Distracting, Not-Worrying and Attention Regulation, to concentrate on aspects of interoception that are likely to show meaningful differences, thereby minimising potential noise. Emotional Awareness, Self-Regulation, Body Listening and Trusting subscales were excluded as they emphasise emotional and therapeutic dimensions of interoception that are less applicable in a sport-specific context such as running. To ensure contextual relevance, all questionnaires were slightly adapted to reflect participants' experiences during running. Rather than assessing an individual's general interoceptive tendencies, the questions were tailored specifically to the run they had just completed. Additionally, while the original questionnaire used a frequency-based scale (ranging from Never to Always), we replaced this with a Likert scale ranging from Strongly Disagree to Strongly Agree, for the same contextual reasons. The fully adapted questionnaire is provided in Appendix B. A completed questionnaire yields a final score between 1 (lowest possible score) and 5 (highest possible score).

Furthermore, we are interested in whether the nature of participants' reflections, while reviewing their run data, differs between the two groups. This will be examined qualitatively through a think-aloud protocol conducted after the run. During this phase, participants are asked to verbally express their thoughts while reviewing their data on the smartphone. This session is split up into two parts, in which the first part will be unprovoked, meaning that the participants are not nudged to talk about anything in specific. In the second part, the participant is asked some specific questions in relation to the data. This second part is added because we anticipated that participants might have very diverse reflections, which makes coding the transcripts more difficult. This could be mitigated by asking all participants the same set of questions

Additionally, the value, experience and usability of the intervention will be explored through semi-structured interviews, conducted exclusively with participants from the intervention group. The think aloud and interview questions can be found in Appendix C. See Figure 13 for a visualisation of the study design.

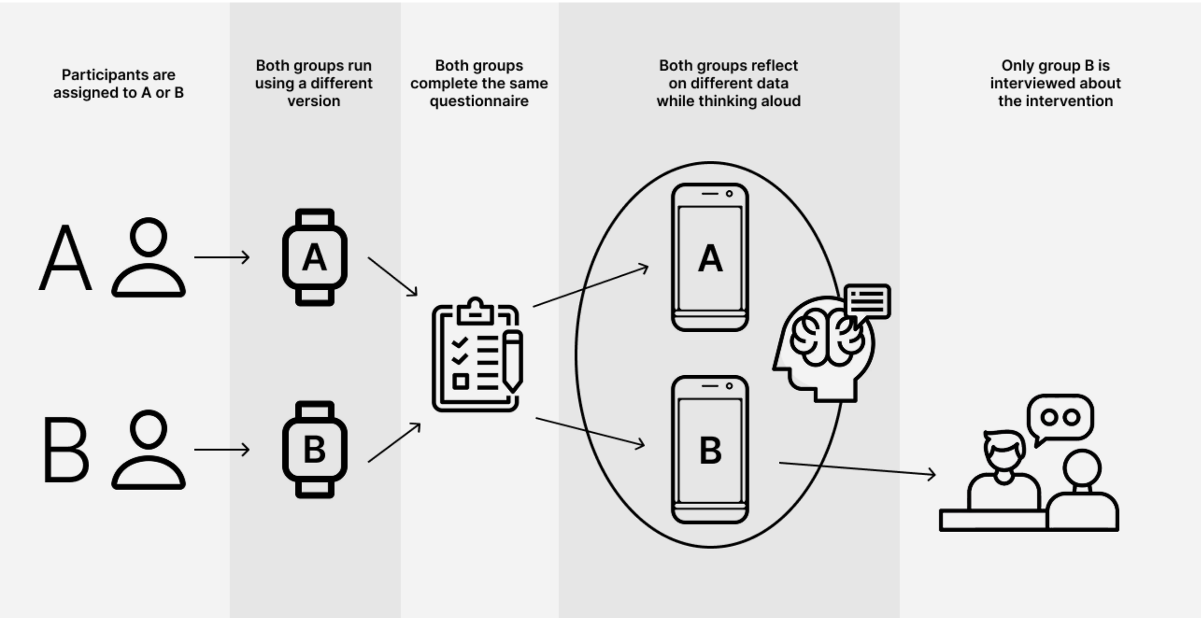


Figure 13
Study design

4.1.1 Conditions

Participants in this study are randomly assigned to either Group A (baseline) or Group B (intervention). The Group B will receive the intervention as presented in section 3.4. In contrast, Group A is intended to have an experience with an ordinary run tracking system. Naturally, Group A will not be prompted with the subjective questions during the run and therefore there is also no post-run reflection on any subjective aspect. However, the typical WBFP-scale including smiles as often seen in self-tracking apps is included. The designs for the baseline can be found in Appendix D and are based on the running apps analysed in section 2.1. The key differences between Group A and Group B are shown in the Table 4 below.

Table 4
Intervention conditions

	Group A - Baseline	Group B - Intervention
Smartwatch during run	On demand interaction with available data: - Pace, - BPM, - Elapsed time - Distance covered.	On demand interaction with available data: - Elapsed time - Distance covered. Periodically prompt including - Reflection point - Subjective question. - Heart rate data
Post-run Reflection on smartphone	- Training summary - Map including marker with pace per kilometre - Heart rate graph	- Training summary - Map including markers with subjective data and heart rate data per kilometre - Body awareness advice based on subjective data

4.1.2 Materials

The core of the intervention was implemented on a **Samsung Galaxy Watch 6**, made available by the Human Media Interaction (HMI) group at the University of Twente. This smartwatch ran a custom-built application developed in **Kotlin** for Wear OS, which implemented the full logic of the intervention, including timing, prompt delivery, and data collection.

For the post-run reflection, participants reviewed their run data using the researcher's **Samsung Galaxy S21 smartphone**. This phone accessed a web-hosted reflection application developed in **Vue.js**, which presented the recorded data back to the participant.

To quantitatively assess interoceptive awareness, participants completed a questionnaire which was implemented using **Google Forms,** allowing for easy data transformation for later analysis[9]. The qualitative components of the study, including a think-aloud script and a set of semi-structured interview questions, were printed on paper. Verbal data from the think-aloud sessions and interviews were recorded using **Voice Memos** on a **MacBook Air**.

4.1.3 Participant Recruitment

The scope of this evaluation was limited to two full-time workweeks, within which all data collection had been completed. Each session with a participant was estimated to take approximately 1.5 to 2 hours, including onboarding, running, questionnaire completion, and participation in a semi-structured interview. Based on this timeframe and available resources, we aimed to recruit a total of 30 participants, with 15 participants assigned to each of the two groups. Participants were randomly assigned to either the baseline or the intervention to reduce allocation bias. Participants were recruited using convenience sampling, primarily through local sports clubs, university networks, and social media channels. This method was chosen due to time constraints and logistical feasibility, although it may limit the generalizability of the results.

[9] https://docs.google.com/forms

To ensure safety and relevance to the nature of the intervention (which involves physical exertion), we only included participants aged 18-65 years. This age range was chosen to focus on the general adult population while avoiding minors and potential risks associated age-related physical limitations. In addition, participants were required to engage in physical activity (specifically running or similar aerobic exercise) at least twice a week and to feel comfortable running at least 5 kilometres without difficulty. This criterion was set to ensure a relatively homogeneous level of baseline fitness between participants, allowing for a more reliable comparison of results related to interoceptive awareness, as athletic populations may already have greater body awareness than sedentary individuals (Wallman-Jones et al., 2021).

4.1.4 Data Collection

As preparatory steps, we ensured the participants knew the time, date, and location of the experiment. To increase the show-up rate, the participants were reminded one day before the experiment. The research procedure began in an indoor setting where the participant was welcomed and introduced to the study. Upon arrival, the participant was informed of the study's procedures, followed by a request to carefully read and sign an informed consent form. This document clearly explained the nature of the study, participant rights, confidentiality of data, and any potential risks or benefits associated with participation.

Once consent was obtained, the researcher collected basic demographic information from the participant. After this, the researcher guided them through a short hands-on practice session with the smartwatch. This allowed the participant to interact with the smartwatch, ensuring they were comfortable using the system before starting the actual run. Finally, the researcher explained the running route that the participant would follow and ensured they had all necessary information before the study continued.

Once the onboarding process was complete, the research setting shifted to an outdoor running environment. The researcher did not accompany the participant during the run to maintain the natural experience of the study. At the designated starting point of the run, the researcher activated the intervention on the participant’s smartwatch and the participant began the run.

During the run, the smartwatch automatically collected and recorded various objective measurements. At the same time, the participant responded to the subjective prompts received via the smartwatch. Upon completion of the run, the participant met the researcher at a pre-determined finish point. Here, the researcher stopped the data collection on the smartwatch and removed the device from the participant. The participant then had time to engage in personal post-run activities such as showering and changing clothes, before engaging in the reflection phase. During this time, the researcher processed the data collected by the smartwatch during the run. The data was made available for the participant’s review on a smartphone, providing them with an overview of their performance metrics and responses during the run.

After the participant completed their post-run activities, they were asked to fill out a self-administered questionnaire. After this, the participant was asked to reflect on the captured data using the smartphone. This was done in a think-aloud format which was recorded with the consent of the participant. Finally, a semi-structured interview was held related to the experience and usability of the intervention.

During the debriefing, participants were informed of which group they had been assigned to (baseline or intervention), given the opportunity to ask any questions, and thanked for their participation. Note that this is the procedure for the intervention group. The procedure for the baseline did not include the subjective prompts during the run nor the did the participants get interviewed as a final step.

4.2 Data Analysis

A total of 26 participants were initially recruited for the study. Due to various reasons, 5 participants dropped out, resulting in a final sample of 21 participants who completed the study protocol. Participants were randomly assigned both a unique ID (1–21) and to either A (baseline, n = 10) or B (intervention, n = 11). Due to an unusable audio recording of Participant 7 (assigned to Version B), this individual's think-aloud and interview data were excluded from qualitative analysis, though other data remained part of the study. The sample included 11 male and 10 female participants, with ages ranging from 21 to 57 years. A majority (15 out of 21) reported regular use of a smartwatch prior to the study, suggesting a generally high level of familiarity with wearable fitness technology. See Table 5 below for detailed information on each participant.

Table 5

Participants demographics

	Version	Age	Gender	Experience with Smartwatch
P01	A	25	Female	Never used one
P02	A	26	Male	Currently using
P03	B	25	Male	Currently using
P04	B	27	Female	Stopped using
P05	B	56	Male	Currently using
P06	A	22	Female	Currently using
P07	A	28	Male	Currently using
P08	B	23	Female	Never used one
P09	A	24	Male	Currently using
P10	B	26	Female	Currently using
P11	B	21	Female	Currently using
P12	A	24	Female	Currently using
P13	A	20	Female	Never used one
P14	A	26	Female	Currently using
P15	B	25	Male	Currently using
P16	B	24	Male	Currently using
P17	A	26	Male	Currently using
P18	A	24	Male	Currently using
P19	B	26	Male	Currently using
P20	B	24	Male	Never used one
P21	B	26	Female	Never used one

To analyse the effect of the intervention on interoceptive awareness, we conducted a statistical comparison between two independent groups: a baseline group (n = 10) and an intervention group (n = 11 total). As the final score from the adapted MAIA-2 questionnaire is an ordinal measure derived from a 5-point Likert scale, and given the relatively small sample size, we opted for a non-parametric test. While normality was not formally assessed, the small sample size made it prudent to avoid assuming a normal distribution. Therefore, we used the Mann–Whitney U test to compare the final MAIA-2 scores between the two groups. This test is appropriate for comparing two independent samples when data are ordinal or non-normally distributed and does not require the assumption of homogeneity of variance (Field, 2018). In line with our non-directional hypothesis, since it was unclear whether the intervention would lead to a positive or negative effect, we conducted a two-sided test. Statistical significance was set at $p < .05$, and all analyses were performed using IBM SPSS Statistics.

Both the think-aloud sessions and the semi-structured interviews were audio-recorded to ensure an accurate capture of participants' verbal reflections. Recordings were stored securely on an encrypted hard drive. Following data collection, the recordings were transcribed using TurboScribe, an AI-assisted transcription service [10]. Each transcript was then manually reviewed and corrected to ensure accuracy. Once verified, all transcripts were anonymised, and the original audio recordings were permanently deleted to protect participant privacy. The qualitative data was analysed through a basic thematic grouping approach, in which participant responses were organised into categories based on recurring patterns in the content. This approach allowed insights to emerge from the data without relying on formal coding procedures or a predefined analytical framework.

4.3 Results

This section presents the results of the various measures used in the study. It begins with the quantitative data, including descriptive and inferential statistics generated using SPSS, followed by an overview of the qualitative results.

4.3.1 MAIA-2 Questionnaire

To provide an overview of participant responses in the questionnaire, descriptive statistics were calculated for both the baseline and intervention group. As explained in section 4.1, each completed questionnaire results in a score ranging from 1 to 5 as an indicator of interoceptive awareness. Participants in the baseline group reported an average score of $M = 3.52$, $SD = 0.46$, with a median score of 3.41 and an interquartile range (IQR) of 0.59. Scores ranged from 2.88 to 4.51. In contrast, participants in the intervention group had a higher average score of $M = 3.92$, $SD = 0.39$, with a median of 3.92 and an IQR of 0.58, ranging from 3.21 to 4.51. The distribution of scores for the baseline group showed slight positive skewness (Skewness = 0.95), while the intervention group showed a slight negative skew (Skewness = –0.30).

[10] https://turboscribe.ai/

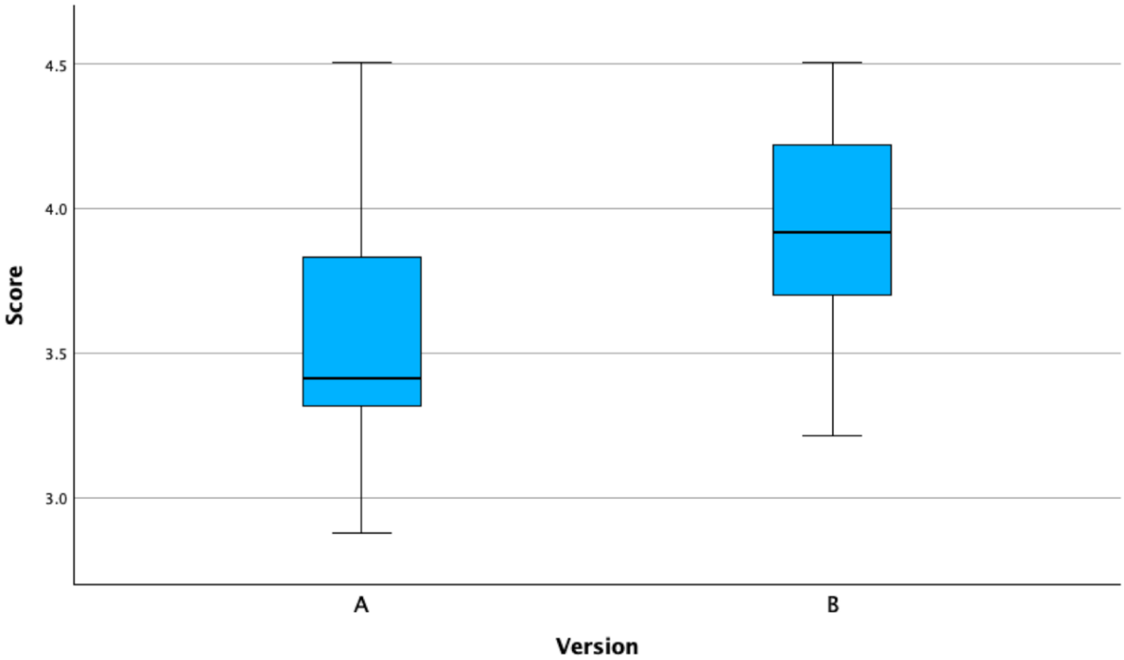


Figure 14
Boxplot quantitative results (A=Baseline; B=Intervention)

To examine whether there was a significant difference in scores between the two groups, a **Mann–Whitney U test** was conducted. Results showed that the questionnaire scores were **significantly higher** in the **intervention group** (*M*-rank = 13.59, *n* = 11) compared to the **baseline group** (*M-r*ank = 8.15, *n* = 10), **U = 26.50**, **Z = –2.008**, **p = .045** (two-tailed). This suggests that participants who received the intervention reported significantly greater interoceptive awareness than those who received the baseline. To calculate the **effect size** for the Mann–Whitney U test, we use the following formula:

$$r = \frac{Z}{\sqrt{N}} = -\frac{2.008}{\sqrt{21}} = -0.44$$

We report the absolute value **r = 0.44**, since the score reflects magnitude, not direction.

4.3.2 Think Aloud

In the unprompted think-aloud sessions, participants in the intervention group, who received subjective reflection prompts during the run, consistently articulated bodily sensations and emotional states while reflecting on their data. Their verbal reports revealed a heightened awareness of internal experiences, often integrating subjective feelings with the objective information provided by the system. Participant 13 described, "*My legs start to feel a bit heavier... but after half a kilometre, it's the same*" while Participant 11 explained, "*If you see the kilometre points, you think about where you were and how you felt about it.*" These reflections suggest that participants were actively interpreting their physiological states in context. In some cases, this involved a direct comparison between felt experience and biometric data. As Participant 21 observed, "*It felt heavier, but the heart rate was similar,*" indicating a conscious evaluation of bodily sensations against the system's feedback.

In contrast, participants in the baseline group, who had access only to objective run data and a basic feeling scale, focused primarily on performance indicators such as pace, distance, and heart rate. Their reflections were largely descriptive and data-driven, with little reference to bodily or emotional states. Participant 24 stated, "*The pace seems to fluctuate… but overall, around average*" while Participant 23 commented, "*I see the map, the kilometres, the heart rate… It's what I expected.*"

During the prompted think-aloud, in which the participants were asked some specific questions, similar patterns emerged. Participants in the baseline group continued to demonstrate reflection on somatic aspects. Participant 8 reflected, "*Even though I felt slower, my heart rate was actually lower. That surprised me*" and Participant 11 remarked, "*I usually don't focus on my belly. Now I did. That's new.*" These statements reflect an openness to bodily perception and a deeper level of reflection.

In contrast, participants in the baseline group provided shorter, less affectively engaged responses when prompted. Their reflections often remained surface-level or returned focus to performance metrics. Participant 24 stated, "I*t was fine. Nice weather. Not much else,"* and Participant 9 said, *"I'll try to keep the same pace next time. Maybe lower heart rate.*" These remarks indicate a more utilitarian orientation, where the reflective process was used primarily to inform performance strategy rather than deepen experiential understanding.

4.3.3 Interviews

The interviews further support the quantitative findings, indicating that the intervention positively influenced participants' interoceptive awareness. Eight out of ten participants reported a noticeable change in their experience during the run. For example, Participant 13 reflected, "*This is more about the feeling that you are looking at… Normally you are always busy with speed and kilometres per minute. And not necessarily with how you feel.*" Similarly, Participant 18 commented on the subjective metric: "*The three options, whether it was lighter, similar or harder, were good for awareness. Often when you just start a normal run yourself, you don't have those kinds of things.*" In terms of behavioural outcomes, four participants also reported making specific adjustments during their run in response to the intervention. These changes ranged from general shifts in attentiveness to concrete adjustments in posture or breathing. As Participant 4 stated, "*I think it is good to just correct your posture. I felt like a professional runner.*"

Participants across the board found the intervention easy to use. This applied both to the during-run component and the post-run reflection. Participant 21 noted, "*Easy, the buttons were clear, it was logical where you had to look.*" Regarding the reflection phase, Participant 18 commented, "*I think it is good to gain even more awareness, instead of just uploading it to Strava and then that is it.*"

However, several challenges and concerns were reported. First, three participants expressed uncertainty about how to relate their subjective bodily sensations to objective performance metrics. For example, participant 21 asked: "*If my legs feel really heavy, is that also a sign that my heart rate is faster? Is that directly related?*" Accordingly, many participants used the prompts to reflect on experiences beyond physical exertion, such as pain, motivation, and decision-making during the run. Participant 18 noted: "*It makes you check in with your body, like, am I really running comfortably?*"

Four participants expressed a desire to access more performance-related data during the run, such as average pace and continuously visible heart rate. While some participants acknowledged that this could undermine the purpose of the reflective prompts, this. Participant 8 remarked, "*I missed that you can actively see your heart rate, but I think that kind of defeats the purpose of that question.*" Similarly, in the post-run reflection, six participants expressed interest in additional performance metrics, including heart rate and speed over time, step frequency, pace per lap, and comparisons to normative data from the general population. However, Participant 16 stated: "I'm of course used to seeing all that data. I personally like that. Maybe it's a bit overkill, though."

Four participants indicated that the on-screen text was initially difficult to read. However, two of them noted that that over time, they no longer needed to read the entire message. Participant 11 shared, “*I was getting the same notifications all the time. So I actually liked that nothing had changed in terms of text or layout. Because that just made me know, oh, the same message.*” Regarding interaction with the watch, two participants experienced difficulty selecting between the “lighter,” “similar,” and “heavier” options. Participant 16 explained, “*You must click a few times every now and then before it really switched.*”

Regarding the perceived value of the intervention, six participants explicitly described the intervention as helpful, primarily citing its ability to foster greater bodily awareness. Participant 11 stated, “I*t helps to adjust your perception of what you feel more with what your body says it is—that you can then feel a little better next time what your heart rate is*.” Furthermore, eight participants expressed willingness to use the intervention again in the future. Participant 11 reflected, “*I would sometimes use it. I think it's fun to think a bit during the run instead of just simply running.*”

Several participants offered suggestions to improve the intervention. Two suggested increasing variation in the types of reflection questions to maintain engagement. As Participant 4 proposed, “*Maybe a few more different things, so that it actually triggers you, instead of you getting the same message.*” Two others expressed a desire for a more nuanced form of body-awareness reflection, potentially allowing for more detailed advice or more diverse input. Additionally, Participant 4 noted that receiving reflection prompts every kilometre may be excessive for longer runs, suggesting a more flexible approach to prompt frequency: “*I think the messages should not be there too often, certainly with longer runs.*” Finally, two participants recommended color-coding the “lighter,” “similar,” and “heavier” response buttons to facilitate easier interaction during movement. As Participant 15 explained, “*The only thing I would change is the colour for the buttons. Then the runner does not even have to look at the text but can just click on the right colour*.”

5. Discussion

5.1 Core Contributions

This study set out to answer the research question: “How can a subjective metric be effectively integrated into self-tracking technology for running such that it enhances the experience of the athlete?” To investigate this, a subjective metric was developed, inspired by DALDA and adapted for conspective, in-situ use. The metric focused on tracking perceived changes in exertion throughout a run, encouraging runners to reflect on their bodily sensations in real time.

Evaluation of the intervention revealed a statistically significant difference in interoceptive awareness scores between group A (control) and group B (intervention), as measured by the MAIA-2 questionnaire. This result is based on the Mann–Whitney U test, which indicated that the difference between the groups was statistically significant. Accordingly, the null hypothesis (H_0): that there is no difference in interoceptive awareness between the groups, was rejected, and we accept the alternative hypothesis (H_1): There is a significant difference in interoceptive awareness scores between the group that received the intervention and the group that did not.

The observed effect size ($r = .44$), interpreted according to Cohen’s (1988) guidelines, reflects a moderate to large effect, suggesting that the intervention had a meaningful impact. These quantitative findings are further supported by the qualitative results. Participants in the intervention group demonstrated more reflective engagement with their feeling during the think-aloud sessions and generally evaluated the intervention positively in post-run interviews, noting its usefulness and expressing a willingness to use it again.

This combination of quantitative and qualitative evidence directly addresses the research question. It demonstrates that a DALDA-inspired conspective subjective metric, focusing on perceived exertion, can be effectively integrated into self-tracking technology for running in a way that enhances interoceptive awareness, thereby positively influencing the runner’s experience.

In the context of existing literature, this work builds on Tholander and Nylander (2015), in which they argue that balancing the measured and lived senses of performance is essential for improving the overall experience of sport. Our findings confirm this, showing that a shift toward introspective awareness through subjective metrics can meaningfully influence how runners engage with their training.

The intervention was also motivated by Karahanoğlu et al. (2024), who emphasise the risks associated with over-reliance on objective feedback in training load management. Karahanoğlu et al. (2024) warn that blind adherence to technology, without considering internal cues, can lead to overtraining, and injury. While our study did not directly assess whether the intervention prevents overtraining or injury, we do see potential for it to contribute to this. This is supported by Wallman-Jones et al. (2021), who define interoception as the brain’s capacity to perceive internal bodily signals such as fatigue, breathing, and heart rate. Interoception is seen as an ability that is essential for regulating effort and avoiding overexertion. If athletes become more attuned to these internal cues during exercise, as our intervention does, they may be better equipped to adjust their activity in real time, potentially mitigating the risks highlighted by Karahanoğlu et al (2024).

Moreover, our findings offer a direction to change the Sport-Data Experience as described by Postma et al. (2024), a concept that captures how athletes subjectively interact with and are affected by sport data. Rather than focusing solely on numerical outputs, the intervention prompted runners to interpret and respond to their bodily sensations during training. In doing so, it illustrates how subjective metrics can actively shape the athlete’s engagement with data and enrich their overall experience, thereby addressing one of the grand challenges in the field of SportHCI (Elvitigala et al., 2024).

Finally, we see potential for the intervention to have an even greater effect in real-world applications, where study-related constraints do not apply. For example, in our study, we deliberately excluded an onboarding component to avoid introducing bias between the baseline and intervention. Although our original plan included onboarding to introduce users to the goal of the intervention and the benefits of improving interoceptive awareness, we omitted this step during the evaluation phase to prevent influencing participant responses. In a real-world commercial context, however, such onboarding would be entirely appropriate and could enhance effectiveness even more. The onboarding designs, which were already completed during development, are included in Appendix E.

5.2 Implications for Design

Drawing from the interview data, several design implications emerge that can guide the development of future iterations. Firstly, the findings highlight the need for a **clear onboarding process**. Although onboarding was excluded in this study to avoid bias, it became apparent that some users would benefit from initial guidance. Specifically, participants were often unsure about how to interpret their subjective sensations in relation to objective heart rate data. Onboarding should clarify the goal of the subjective metric and how it is structured and intended to function alongside performance data.

Secondly, the study revealed that users reflect on a **broad range of sensations** during exercise, not just exertion. While the intervention was designed around exertion-based reflection points, many participants extended their reflections to include discomfort, pain, and emotional states. This suggests that future systems could be designed to accommodate a wider range of subjective metrics, allowing space for psychological and motivational reflections as part of the experience.

Furthermore, the intervention aimed to balance measured and lived performance, as explained by Tholander and Nylander (2015). However, several users expressed a **desire for more performance data**. Suggesting that the intervention leans too far toward the “lived-sense of performance”-side of the spectrum. This can be explained by that people are accustomed to using fitness apps that emphasize objective data over the subjective information, as discussed in Section 2.1. Future systems should aim for flexibility, offering both subjective reflection and performance data.

In addition, the results imply the need for **long-term and context-sensitive adaptation**. The fixed number of prompts used in this evaluation may be suitable for short-term studies but could become excessive or repetitive in day-to-day training contexts. Participants indicated that repeated prompts might become obstructive over time. This aligns with findings by Ben-Nun (2024) mentioned in section 2.2.2, who observed that frequent, unvarying prompts can contribute to user fatigue and reduced engagement over time.

Future systems should consider adaptive strategies that adjust the frequency, timing, and content of prompts based on user preference and training conditions.

Finally, these insights collectively point to the importance of **designing for flexibility**. Different athletes engage with their training and data in different ways. As such, future implementations should allow for personalization in how subjective metrics are introduced, reflected on, and interpreted, ensuring the system remains useful for a wide range of users.

5.3 Limitations and Future Work

While the study yielded promising insights into the integration of subjective metrics in self-tracking technologies, several limitations related to the study design should be acknowledged.

First, the sample size was relatively small (n = 21), limiting the statistical power and generalizability of the findings. A power analysis suggests that, for detecting medium-sized effects (e.g., r = .3) with adequate statistical power (typically 0.8), a larger sample would be required—generally around 64 participants in a between-subjects design. The use of convenience sampling further introduced potential bias, with several participants personally acquainted with the researcher. This may have affected the neutrality of their responses and narrowed the demographic diversity of the sample.

Second, the experimental design involved a single exposure to either the baseline or the intervention, without within-subject comparison. Each participant interacted with the system only once, making it difficult to control for inter-individual differences. According, one participant who was later revealed to have received the baseline noted that he would likely have scored high on the interoceptive awareness measure regardless of the intervention, due to already being highly attuned to his body. Such variation highlights the importance of accounting for baseline introspective tendencies in future studies. Future research could build on these findings by testing a similar intervention with a more robust study design. A within-subject, pre-post setup involving a larger and more demographically diverse participant pool would allow for stronger causal claims and a deeper understanding of how interoceptive awareness evolves with repeated use.

Third, researchers should also explore how such subjective metrics function in long-term, real-world use. Participant feedback in this study suggested that frequent prompting may become disruptive during longer runs. Future work could examine how often such interactions should occur within a structured training plan, and test adaptive prompting strategies that tailor frequency based on user preference or session length. Understanding these dynamics would be crucial for maintaining engagement and ensuring the sustainability of the intervention in practice. This resonated with the open design principles and designing for specific explained in section 3.2.2.

Beyond interoceptive awareness, future research could also investigate how subjective metrics might support other components of the SDX. As discussed in section 2.4, one promising direction is enhancing trust in training load management (TLM) systems. Subjective inputs gathered ex-situ, outside of the training session, may help athletes contextualize automated feedback and improve the perceived reliability of these systems.

Lastly, exploring the broader applicability of subjective metrics across different sports, user groups, and experience levels would be valuable, since it might function differently in various contexts. Whether in cycling, team sports, or among novice athletes, the use of self-reported bodily experience offers potential to enhance engagement, prevent injury, and improve performance across a variety of athletic contexts.

6. Conclusion

This study investigated how a subjective metric could be integrated into self-tracking technology for running in a way that enhances the athlete's experience. Through a review of existing applications and relevant literature, we identified interoceptive awareness as a meaningful focus. We designed a conspective intervention using a smartwatch to prompt athletes to reflect on changes in perceived exertion during the run, supported by post-run feedback comparing these reflections with heart rate data.

The evaluation, conducted with 21 participants, revealed a statistically significant increase in interoceptive awareness in the intervention group, as indicated by the questionnaire. Think-aloud data supported this by showing that participants reflected more on somatic aspects during the run. Interviews further indicated that most participants found the intervention helpful and would consider using it again in future runs, alongside other design implications.

While the scope of this study was limited due to time constraints, resulting in a sub-optimal study design and short-term intervention testing, the findings suggest that integrating subjective metrics during activity can positively influence how runners engage with their bodily sensations. These results highlight the potential of subjective metrics to enrich the Sport-Data Experience and provide useful insights for the design of self-tracking systems that move beyond purely performance-oriented feedback.

References

Ahmad, K. S., Ahmad, N., Tahir, H., & Khan, S. (2017). Fuzzy_MoSCoW: A fuzzy based MoSCoW method for the prioritization of software requirements. 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). https://doi.org/10.1109/icicict1.2017.8342602

Ben-Nun, P. (2024). *Respondent Fatigue*. Sagepub.com. https://methods.sagepub.com/ency/edvol/encyclopedia-of-survey-research-methods/chpt/respondent-fatigue

Bentvelzen, M., Niess, J., & Woźniak, P. W. (2022). Designing Reflective Derived Metrics for Fitness Trackers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *6*(4), 1–19. https://doi.org/10.1145/3569475

Calvert, T. W., Banister, E. W., Savage, M. V., & Bach, T. (1976). A Systems Model of the Effects of Training on Physical Performance. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-6*(2), 94–102. https://doi.org/10.1109/tsmc.1976.5409179

Casey, A., & Kirk, D. (2020). *Models based Practice in Physical Education*. https://doi.org/10.4324/9780429319259

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Journal of the American Statistical Association*, *73*(363), 680. https://doi.org/10.2307/2286629

Delden, V., Moreno, A., Poppe, R., Reidsma, D., & Heylen, D. (2014). *Steering Gameplay Behavior in the Interactive Tag Playground*. https://doi.org/10.1007/9783319141121

Den Hartigh, Ruud. J. R., Meerhoff, L. R. A., Van Yperen, N. W., Neumann, N. D., Brauers, J. J., Frencken, W. G. P., Emerencia, A., Hill, Y., Platvoet, S., Atzmueller, M., Lemmink, K. A. P. M., & Brink, M. S. (2022). Resilience in sports: a multidisciplinary, dynamic, and personalized perspective. *International Review of Sport and Exercise Psychology*, *17*(1), 1–23. https://doi.org/10.1080/1750984x.2022.2039749

Desmedt, O., Olivier Luminet, Maurage, P., & Corneille, O. (2023). Discrepancies in the Definition and Measurement of Human Interoception: A Comprehensive Discussion and Suggested Ways Forward. *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916231191537

Doherty, K., & Doherty, G. (2018). The construal of experience in HCI: Understanding self-reports. *International Journal of Human-Computer Studies*, *110*, 63–74. https://doi.org/10.1016/j.ijhcs.2017.10.006

Edwards, W. H. (2010). *Motor Learning and Control: From Theory to Practice*. Cengage Learning.

Elvitigala, D., Karahanoglu, A., Matviienko, A., Vidal, T., Postma, D., Jones, M., Montoya, M., Harrison, D., Elbæk, L., Daiber, F., Burr, L., Patibanda, R., Buono, P., Hämäläinen, P., Delden, V., Bernhaupt, R., Ren, X., Rheden, van, Zambetta, F., & Mueller, F. (2024). *Grand Challenges in SportsHCI*. https://doi.org/10.1145/3613904.3642050

Eston, R. (2012). Use of Ratings of Perceived Exertion in Sports. *International Journal of Sports Physiology and Performance*, *7*, 175–182. https://doi.org/10.1123/ijspp.7.2.175

Feng, S., Mäntymäki, M., Dhir, A., & Salmela, H. (2020). How Self-tracking and the Quantified Self Promote Health and Well-being: A Systematic Literature Review (Preprint). *Journal of Medical Internet Research*, *23*(9). https://doi.org/10.2196/25171

Field, A. P. (2018). Discovering Statistics Using IBM SPSS Statistics. 5th Edition, Sage, Newbury Park. - References - Scientific Research Publishing. Www.scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=3504991

Grant, S., Aitchison, T., Henderson, E., Christie, J., Zare, S., Mc Murray, J., & Dargie, H. (1999). A Comparison of the Reproducibility and the Sensitivity to Change of Visual Analogue Scales, Borg Scales, and Likert Scales in Normal Subjects During Submaximal Exercise. *Chest*, *116*(5), 1208–1217. https://doi.org/10.1378/chest.116.5.1208

Hortigüela Alcalá, D., & Hernando Garijo, A. (2017). Teaching Games for Understanding: A Comprehensive Approach to Promote Student's Motivation in Physical Education. *Journal of Human Kinetics*, *59*(1), 17–27. https://doi.org/10.1515/hukin-2017-0144

Ishii, H., Wisneski, C., Orbanes, J., Chun, B., & Paradiso, J. (1999). *PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play*. 394–401. https://doi.org/10.1145/302979.303115

Karahanoğlu, A., Coskun, A., Postma, D., Leonard Scheltinga, B., Gouveia, R., Reidsma, D., & Reenalda, J. (2024). *Is it just a score? Understanding Training Load Management Practices Beyond Sports Tracking*. https://doi.org/10.1145/3613904.3642051

Karahanoğlu, A., Gouveia, R., Reenalda, J., & Ludden, G. (2021). How Are Sports-Trackers Used by Runners? Running-Related Data, Personal Goals, and Self-Tracking in Running. *Sensors*, *21*(11), 3687. https://doi.org/10.3390/s21113687

Kellmann, M., & Kallus, W. (2001). *The recovery-stress-questionnaire for athletes* (pp. 76–136). Human Kinetics.

Khatri, A., & Kalra, N. (2012). A Comparison of Two Pain Scales in the Assessment of Dental Pain in East Delhi Children. *ISRN Dentistry*, *2012*, 247351. https://doi.org/10.5402/2012/247351

Loerakker, M. B., Niess, J., Bentvelzen, M., & Woźniak, P. W. (2023). Designing Data Visualisations for Self-Compassion in Personal Informatics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *7*(4), 1–22. https://doi.org/10.1145/3631448

Márquez Segura, E., Turmo Vidal, L., & Rostami, A. (2016). Bodystorming for movement-based interaction design. *Human Technology*, *12*(2), 193–251. https://doi.org/10.17011/ht/urn.201611174655

Mehling, W., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The Multidimensional Assessment of Interoceptive Awareness, Version 2 (MAIA2). *PLOS ONE*, *13*, e0208034. https://doi.org/10.1371/journal.pone.0208034

Newell, K., & Jordan, K. (2007). *Task Constraints and Movement Organization: A Common Language*. https://doi.org/10.5040/9781492595434.ch001

Niess, J., Knaving, K., Kolb, A., & Woźniak, P. W. (2020). *Exploring Fitness Tracker Visualisations to Avoid Rumination*. https://doi.org/10.1145/3379503.3405662

Noakes, T. D. (2008). Testing for maximum oxygen consumption has produced a brainless model of human exercise performance. *British Journal of Sports Medicine*, *42*(7), 551–555. https://doi.org/10.1136/bjsm.2008.046821

Palsa, L., & Mertala, P. (2023). Contextualizing Everyday Data Literacies: The Case of Recreational Runners. *International Journal of Human-Computer Interaction*, 1–12. https://doi.org/10.1080/10447318.2023.2241612

Pires, F. O., Noakes, T. D., Lima-Silva, A. E., Bertuzzi, R., Ugrinowitsch, C., Lira, F. S., & Kiss, M. A. P. D. M. (2011). Cardiopulmonary, blood metabolite and rating of perceived exertion responses to constant exercises performed at different intensities until exhaustion. *British Journal of Sports Medicine*, *45*(14), 1119–1125. https://doi.org/10.1136/bjsm.2010.079087

Postma, D., Delden, V., Koekoek, J., Walinga, W., Hilvoorde, I., BEIJNUM, B., Salim, F., & Reidsma, D. (2022). *A Design Space of Sports Interaction Technology*. https://doi.org/10.1561/9781638280651

Postma, D., Reidsma, D., van Delden, R., & Karahanoğlu, A. (2024). From Metrics to Experiences: Investigating How Sport Data Shapes the Social Context, Self-Determination and Motivation of Athletes. *Interacting with Computers*. https://doi.org/10.1093/iwc/iwae012

Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. *Medical & Biological Engineering & Computing*, *44*(12), 1031–1051. https://doi.org/10.1007/s11517-006-0119-0

Ralph, R. (1993). A review of the profile of mood states (POMS) in the prediction of athletic success. *Journal of Applied Sport Psychology*, *5*(1), 78–84. https://doi.org/10.1080/10413209308411306

Rapp, A., & Tirabeni, L. (2020). Self-tracking while doing sport: Comfort, motivation, attention and lifestyle of athletes using personal informatics tools. *International Journal of Human-Computer Studies*, *140*, 102434. https://doi.org/10.1016/j.ijhcs.2020.102434

Renshaw, I., Davids, K., Araújo, D., Lucas, A., Roberts, W. M., Newcombe, D. J., & Franks, B. (2019). Evaluating Weaknesses of “Perceptual-Cognitive Training” and “Brain Training” Methods in Sport: An Ecological Dynamics Critique. *Frontiers in Psychology*, *9*(2468). https://doi.org/10.3389/fpsyg.2018.02468

Ritchie, C. (2012). Rating of Perceived Exertion (RPE). *Journal of Physiotherapy*, *58*(1), 62. https://doi.org/10.1016/s1836-9553(12)70078-4

Rolstad, S., Adler, J., & Rydén, A. (2011). Response Burden and Questionnaire Length: Is Shorter Better? a Review and Meta-analysis. *Value in Health*, *14*(8), 1101–1108. https://doi.org/10.1016/j.jval.2011.06.003

Saw, A., Main, L., & Gastin, P. (2015). Monitoring the athlete training response: Subjective selfreported measures trump commonly used objective measures: A systematic review. *British Journal of Sports Medicine*, *50*. https://doi.org/10.1136/bjsports2015094758

Scheid, J. L., & O’Donnell, E. (2019). REVISITING HEART RATE TARGET ZONES THROUGH THE LENS OF WEARABLE TECHNOLOGY. *ACSM’s Health & Fitness Journal*, *23*(3), 21–26. https://doi.org/10.1249/fit.0000000000000477

Schell, J., & Schell, J. (2008). *The Art of Game Design*. https://doi.org/10.1201/9780080919171
Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00395
Swart, J., P, L. R., I, L. M., V, L. E., W, W. R., Johnston, S., & D, N. T. (2009). Exercising with reserve: exercise regulation by perceived exertion in relation to duration of exercise and knowledge of endpoint. *Br J Sports Med*, *43*(10), 775. https://doi.org/10.1136/bjsm.2008.056036
Tholander, J., & Nylander, S. (2015). Snot, Sweat, Pain, Mud, and Snow. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2702123.2702482
VanGundy, A. B. (1984). BRAIN WRITING FOR NEW PRODUCT IDEAS: AN ALTERNATIVE TO BRAINSTORMING. *Journal of Consumer Marketing*, *1*(2), 67–74. https://doi.org/10.1108/eb008097
Wallman-Jones, A., Perakakis, P., Tsakiris, M., & Schmidt, M. (2021). Physical activity and interoceptive processing: Theoretical considerations for future research. *International Journal of Psychophysiology*, *166*, 38–49. https://doi.org/10.1016/j.ijpsycho.2021.05.002
Weinberg, R. S., & Gould, D. (2023). *Foundations of Sport and Exercise Psychology* (8th ed.). Human Kinetics.
Williams, N. (2017). The Borg Rating of Perceived Exertion (RPE) Scale. *Occupational Medicine*, *67*(5), 404–405. https://doi.org/10.1093/occmed/kqx063
Zinoubi, B., Zbidi, S., Vandewalle, H., Chamari, K., & Driss, T. (2017). Relationships between rating of perceived exertion, heart rate and blood lactate during continuous and alternated-intensity cycling exercises. *Biology of Sport*, *35*(1). https://doi.org/10.5114/biolsport.2018.70749

Appendix A

Ethical Approval

UNIVERSITY OF TWENTE.

Computer & Information Sciences (CIS)

250715 APPLICATION FOR ETHICAL REVIEW

Application nr:	250715	**Intro form:**	8 - Introduction
Researcher:	van Mensvoort, W.S.H.M-ITECH)	**Middle form:**	11 - Computer & Information Science (CIS)
Supervisor:	Reidsma, D. (EEMCS-HMI)	**Outro form:**	5 - Submission
Reviewer:	van der Ham - de Vos, J. (EEMCS-DACS)		
Status:	Positive advice by reviewer		
Date of application:	09-04-2025 14:05		
Application version:	2		

0. GENERAL

0.1. Personal details

Student/employee number: s3072045
Initials: W.S.H.
First name: Wout
Last name: van Mensvoort
Email : w.s.h.vanmensvoort@student.utwente.nl
Education/department: n/a
Faculty: n/a
Study field: M-ITECH
Study level: MSC
Faculty/service department: EEMCS (Selected for this application)

0.2. Project title

Rethinking Subjective Metrics in Self-Tracking Systems for Running

1 / 9

1.2. PRIVACY, GDPR, AND POSSIBLE NEED FOR DPIA: Does the research include any possible access to, gathering, or use, or publication of data that can be traced back to specific individuals, directly or, for instance, by combining data from multiple sources? Or is it possible that you will accidentally access or publish Personal Identifiable Information (PII)?

Yes, and we follow the rules on processing of personal data, including acquiring explicit consent for processing PII (besides possibly the consent for participating in research) and including a possibly necessary GDPR registration.

1.3. RESEARCH DOMAIN: Regarding the nature of your research, does one or more of the following statements apply to your research?
the research is in a potentially medical domain such as illness, assessment and diagnosis, prevention, cure, or carethe research addresses a health outcomethe research gathers health datathe research involves a hospital or other medical settingthe research may be potentially medical for some (other) reason

No, the research is not medical, health related, or close-to-medical in any way whatsoever

2. HUMAN RESEARCH PARTICIPANTS

2.1. HUMAN PARTICIPANT RESEARCH: Does the research include
a) active involvement of human research participants during the research, and/or
b) gathering new data from individuals
such as measurements or responses from interviewees, survey respondents, participants, informants, or simply people whose data is measured because they are present in a certain place at a certain time?

Yes, my research falls under "research with human participants'

3. RESEARCH POPULATION

3.1. RESEARCH POPULATION: Please provide a brief description of the intended research population, including inclusion and exclusion criteria, number of participants, and recruitment strategies.

The study will involve approximately 20 participants. Inclusion criteria include are between 18 and 65 years old, having no history of cardiovascular disease, and being able to comfortably complete a 5K run. Participants should also engage in regular physical activity (at least twice per week) and be willing to wear a smartwatch during the run while responding to subjective prompts. Recruitment will be done through convenience sampling, using channels such as university mailing lists, social media, and personal networks.

3.2. LACK OF CAPACITY TO CONSENT: Do you have participants who are formally NOT able to give informed consent?

No, all participants have the capacity to consent

3.3. VULNERABLE PARTICIPANTS: Does your research target vulnerable participants such as focusing on specific ethnic groups, people in another country, minor (<16 years), people with physical or cognitive impairments (regardless of their capacity to consent), people under institutional care (e.g., nursing homes, hospitals, prisons), or any other particular group that may be more vulnerable than other people in the general population?

No

3.4. POWER RELATIONS: Does your research target participants somehow dependent on, or in a subordinate position to the researcher (e.g., students or relatives)?

No

4. RESEARCH PROCEDURE AND RISKS

4.1. RESEARCH TYPES: Which of the following research types do you employ in your research?

Interviewing and surveys: paper/online questionnaires, survey, face to face or online interview, focus group

Participating in non-experiment activity: can be formative evaluation of prototypes, but more in general providing artificial tasks, including triggering stimuli and tasks to elicit observable behavior and responses; measured with e.g. observations, interviews, and manual or automated data collection

4.2. CONTEXT OF REAL LIFE ACTIVITIES: Do the activities of participants that people do, included in your research, include activities in a real life setting?

Yes

4.3. MATERIALS, PROTOTYPES AND DESIGNS: Do the activities include interaction with a prototype, design, mockup, product, interaction technology, etc?

Yes

4.4. ASSIGNING TASKS TO PARTICIPANTS: Do the research procedures include activities performed specifically for the sake of the research?

Yes

4.5. LOCATION: Where will the research activity take place?

The research will take place onsite at The University of Twente. The part of the research that takes place before and after the run will be inside. Public places like Bastille, Vrijhof or sports canteen are possible locations, but this is depending on how crowed these places are on the day of research. The running part of the activity will be outside. A track will be laid out for the participants to run.

4.6. TIME INVESTMENT: How much time will each participant spend?

With each participant I will meet up once and I estimate a session to take 120 minutes.

4.7. DESCRIPTION OF RESEARCH PROCEDURE: What is the research procedure, in terms of setting, tasks, activities, content, and stimuli?

The research procedure will begin at an indoor location where the participant will be welcomed and introduced to the study. At this point, they will be asked to carefully read and sign an informed consent form. Once consent is obtained, the researcher will provide a briefing outlining the purpose and protocol of the study. The aim of the study is explained and the importance of accurately assessing physical exertion is stressed. It is explained that this ability is valuable for preventing injuries and overtraining, as well as for helping individuals safely push themselves during peak moments, such as the final stretch of a race.

Following the introduction, the participant will receive detailed information about the intervention and the subjective metrics they will be asked to assess. This includes an explanation of the types of questions they will be answering, the frequency and timing of these prompts, how responses should be given, and how the smartwatch will notify them to respond. Participants will also have a chance to practice interacting with the smartwatch to ensure they are comfortable with the system prior to the run. The onboarding for group A will be slightly different, since they will not be answering the subjective questions during the run.

Furthermore, the participant will be informed about the specific route they will have to run.

After the onboarding process is complete, the research setting will shift to an outdoor running environment. Although the researcher will not accompany the participant during the run (to preserve the natural feel of the experience), this could be reconsidered for logistical reasons if necessary.

At the starting point of the run, the researcher will initiate the intervention on the participant's smartwatch. The participant will then begin the run and, at designated intervals, will receive prompts from the smartwatch.

Upon completion of the run, the participant will meet the researcher at the pre-determined finish point, where the researcher will stop the data collection on the smartwatch. The participant will then have time for any personal post-run activities, such as showering and changing clothing.

During this time, the researcher will process the data collected during the run and make it available for review on a smartphone.

The participant will then be asked to complete a questionnaire regarding their interoceptive awareness during the run.

Finally, the participants will do a think aloud reflection on their captured data, followed by an more in-dept interview exploring topics such as whether the intervention enhanced their awareness of bodily sensations, their overall experience, and the usability of the system.

4.8. MEASURES: What measurements, recording tools, discussion topics will you employ?

During the run, the smartwatch will be used to collect and record data. The smartwatch will continuously be tracking heart rate and gps data (automated measurements). The smartwatch will also be used to input the subjective metric (self report). During the run the researcher will not be present. After the run, (where is researcher is present), the participants will take a self-administered questionnaire. Finally, the participant is asked to reflect on their recorded data while thinking out loud. Followed by an interview about their experiences with the intervention. An audio recording will be made of the reflection and the interview which will be transcribed for future analysis. The participants are made aware of this recording and how we will safely handle this recorded-file.

4.9. RISK OF ADVERSE EFFECTS: Is there a risk for adverse (or: negative) effects of the research for certain participants, and how do you deal with these risks?

No

4.10. BURDEN TO THE PARTICIPANT: Are there other short-term or long-term burdens and/or risks to the participants?

No

4.11. ACCIDENTAL FINDINGS: Does the method used allow for making an accidental, diagnostic finding that the experimental participant might have to be informed about?

No, the method does not allow for this possibility

4.12. COVID19: Are you aware of departmental/UT rules regarding experimentation under COVID19 and will you follow them?

Yes I know the rules and will follow them
No rules apply

5. (DE-)BRIEFING, DECEPTION & CONSENT PROCEDURE

5.1. BRIEFING. Will you inform potential research participants (and/or their legal representatives in case of legally non-competent participants) completely about the aims, activities, burdens and risks (such as to their health and well-being) of the research and about other relevant information before the decide to take part in the research? How will you do this?

Yes, participants are fully briefed beforehand

5.2. Please explain

No information will be withhold

5.3. If applicable, upload your information letter as a PDF

Information letter (1).pdf

5.4. INFORMATION ON WITHDRAWAL OF CONSENT. Will you inform potential research participants (and/or their legal representatives in case of legally non-competent participants) clearly that they can withdraw from the research at any time without explanation/justification?

Yes

5.5. DECEPTION. Will you use any Deception in the research procedure? How, and why?

No, we will not use any deception

5.6. DEBRIEFING: Will the research procedure involve a debriefing after participation, and how will you do this?

Yes

The participant will be informed if he/se tested the baseline or the intervention. Also there is room for the participant to ask questions about the study.

5.7. FREEDOM TO PARTICIPATE: Are the participants completely free to participate in the research and to withdraw from participation whenever they wish and for whatever reason?

Yes, and we clearly communicate this to them

5.8. DIRECT CONSENT OR PROXY CONSENT: Who will provide the consent?

Participant (no legal representative will separately be informed)

5.9. TYPE OF CONSENT: Which type of consent will you use?

Signed, written consent form prior to participation

5.10. Please upload supportive information

Informed Consent.docx (1).pdf

5.11. CONSENT FOR FUTURE USE: Will you keep and reuse the newly collected data for future research, and do you obtain adequate consent for this?

No, I will only use the data for this research

5.12. PERSONAL DATA: Will you gather new personally identifiable data, about the research participants, and does the consent information also address consent for Personally Indentifiable Information (PII), separate from and in addition to consent for research participation and research data collection and use?

Yes, I will gather PII; the consent information deals with this separately and adequately

5.13. PUBLICATION OF THE DATA: Will you publish (some of) the newly collected data, and do you obtain adequate consent for this?

Yes, I will make (some of) the data publicly available, and I do obtain explicit consent

5.14. REWARDS: Will participants receive any rewards, incentives or payments for participating in the research?

No, the participants will not receive any reward

6. PRE-EXISTING DATA

6.1. PRE-EXISTING DATA: Will the research involve the inclusion, combination, use, and/or analysis of already existing data sets about people?

No

7. AI TECHNOLOGY

7.1. AI TECHNOLOGY: Will the project develop AI technology, or will the project involve the deployment and/or use of AI technology for practical applications?

No

8. CYBERSECURITY

8.1. CYBERSECURITY: Will the research involve any cybersecurity or online privacy issues, such as the possible discovery of security vulnerabilities, experiments with malicious software (e.g., computer viruses), or the discovery and investigation of illegal activities on the Internet?

No

9. UNINTENDED CONSEQUENCES, MISUSE, AND APPLICATION RISKS

9.1. MISUSE: Is it reasonable to anticipate that the research will provide knowledge, products, or technologies that could be intentionally used to threaten, or non-intentionally result in threats, to public health and safety, crops and other plants, animals, the environment, or material infrastructure?

No

9.2. INCLUSIVITY AND SOCIAL INJUSTICE: Is a disproportionally negative impact foreseeable on certain groups of users or non-users, for example, people of a certain age, gender, sexual orientation, social class, race, ethnicity, religion, political orientation, culture, or disability, creating or reinforcing social injustices?

No

9.3. MILITARY APPLICATION: Does your research or prototype have military/police/defense applications?

No

10. OTHER ETHICAL ISSUES

10.1. CONFLICTS OF INTEREST: Do any of the parties involved in overseeing or carrying out the research have a potential conflict of interest?

No

10.2. RISKS TO THE RESEARCHER: Will the study expose the researcher to any risks (e.g. when collecting data in potentially dangerous environments or through dangerous activities, when dealing with sensitive or distressing topics, or when working in a setting that may pose 'lone worker' risks)?

No

10.3. OTHER POTENTIAL ETHICAL ISSUES: Do you anticipate any other ethical issues in your research project that have not been previously noted in this application?

No

11. CLOSURE

11.1. I have answered all questions truthful and complete

Yes

12. COMMENTS

Reidsma, D. (EEMCS-HMI) (09-04-2025 14:20) :

5.12 and 1.2 should be yes (you DO collet pii, namely at least the audio, and should properly explain on concent form, ask explicit permission for that, and show how you store safely (encrypted HDD) and retain it for minimum possible time (transcribe, then immedaitely throw away audio)

13. CONCLUSION

Status: Positive advice by reviewer

15-04-2025 11:00

Appendix B

Questionnaire

Noticing
Q1: During the run, I noticed where tension was located in my body.
Q2: During the run, I noticed when I felt physically uncomfortable.
Q3: During the run, I noticed where in my body I felt comfortable.
Q4: During the run, I noticed changes in my breathing, like whether it sped up or slowed down.

Not-Distracting
*Q5: During the run, I ignored physical tension or discomfort until it became more severe.
*Q6: During the run, I distracted myself from sensations of discomfort.
*Q7: During the run, I tried to power through pain or discomfort.
*Q8: During the run, I tried to ignore any pain I was feeling.
*Q9: During the run, I pushed feelings of discomfort away by focusing on something else.
*Q10: During the run, I occupied myself with other thoughts to avoid feeling unpleasant sensations.

Not-Worrying
*Q11: During the run, feeling physical pain made me feel upset.
*Q12: During the run, I started to worry that something was wrong when I felt discomfort.
Q13: During the run, I noticed unpleasant sensations without getting worried.
Q14: During the run, I stayed calm even when I felt pain or discomfort.
*Q15: During the run, I couldn't get feelings of discomfort or pain out of my mind.

Attention Regulation
Q16: During the run, I could focus on my breath even if things were happening around me.
Q17: During the run, I maintained awareness of my bodily sensations even with distractions.
Q18: During the run, I could stay aware of my posture.
Q19: During the run, I was able to return attention to my body when I got distracted.
Q20: During the run, I could shift from thinking to sensing my body.
Q21: During the run, I maintained awareness of my whole body, even when part of it felt discomfort.
Q22: During the run, I was able to consciously focus on my body as a whole.

The statements were answered using a 5-Likert scale ranging from "Strongly disagree" to "Strongly Agree". For the questions marked with an * the scoring is inversed.

The result of the questionnaire is shown the table below. The rows represent participants, and the columns represent the questions.

		Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14	Q 15	Q 16	Q 17	Q 18	Q 19	Q 20	Q 21	Q 22
A	P 1	5	5	5	4	3	2	2	3	1	2	4	4	4	4	4	5	4	5	5	5	5	3
A	P 2	3	2	4	5	2	4	4	4	4	4	4	4	2	4	2	4	4	4	3	4	4	4
B	P 3	5	4	5	4	1	2	1	3	3	4	5	4	4	5	4	5	5	5	4	5	5	5
B	P 4	4	4	4	4	3	2	2	3	1	3	4	2	5	5	3	3	4	2	3	3	2	2
B	P 5	3	5	5	4	1	2	1	2	1	1	4	5	5	5	4	4	4	5	4	5	4	5
A	P 6	3	4	4	5	2	2	1	2	1	2	3	4	4	4	4	4	2	3	4	2	2	3
A	P 7	4	4	3	4	2	2	2	4	4	5	2	2	4	5	4	4	3	3	3	4	1	3
B	P 8	4	4	4	4	4	2	1	4	3	3	5	5	4	5	4	4	2	2	4	5	4	4
A	P 9	4	5	4	3	2	1	2	2	2	2	4	5	4	5	3	4	5	4	4	4	3	3
B	P 10	5	5	4	4	2	2	2	4	2	2	4	5	4	4	2	4	4	3	4	4	2	2
B	P 11	5	4	5	5	4	5	2	4	5	5	3	5	5	5	4	5	5	5	3	5	5	5
A	P 12	4	4	4	4	2	4	4	4	3	4	5	5	4	4	4	3	3	2	4	4	4	4
A	P 13	5	4	5	5	4	5	2	4	5	5	3	5	5	5	4	5	5	5	3	5	5	5
A	P 14	3	4	2	5	5	4	4	4	2	2	4	2	2	4	5	4	2	4	4	2	2	2
B	P 15	5	4	4	5	4	5	4	4	5	5	4	5	1	5	4	4	4	5	5	4	4	5
B	P 16	4	4	5	5	3	4	5	4	5	5	4	5	4	4	5	5	4	3	3	4	4	3
A	P 17	3	4	2	2	4	5	4	4	4	1	4	5	2	4	5	4	4	1	3	3	4	2
A	P 18	1	4	1	2	5	2	2	1	1	4	3	2	2	4	4	5	4	4	4	4	4	3
B	P 19	4	4	4	4	3	2	2	4	3	4	4	5	2	3	5	5	5	5	4	4	4	4
B	P 20	5	4	5	4	1	4	2	4	2	2	5	5	5	5	4	5	5	5	5	5	5	5
B	P 21	4	4	4	3	4	3	5	4	4	4	5	2	4	5	5	3	4	4	4	4	4	3

Appendix C

Think-aloud and Interview

Think aloud
Unprovoked

No questions asked

Provoked
1. How did you feel during the run?

2. Is there anything in the data that stands out to you?

3. How would you use this data for your next run?

Interview
1. Can you describe your overall experience using the intervention during the run?

2. How easy or difficult was it to use the intervention?

3. Did you encounter any problems or barriers during the use?

4. In what way did the intervention change your running experience in comparison to running without it?

5. Can you describe your experience with the post-run reflection?

6. What did you like about the reflection and what did you not like?

7. Do you think this reflection could influence your next run? If yes how

8. Did you feel the intervention was helpful? In what ways?

9. Would you use it again? Why or why not?

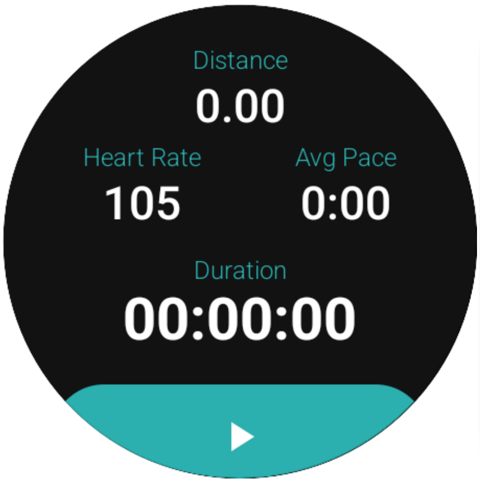
10. What would you improve about the intervention?

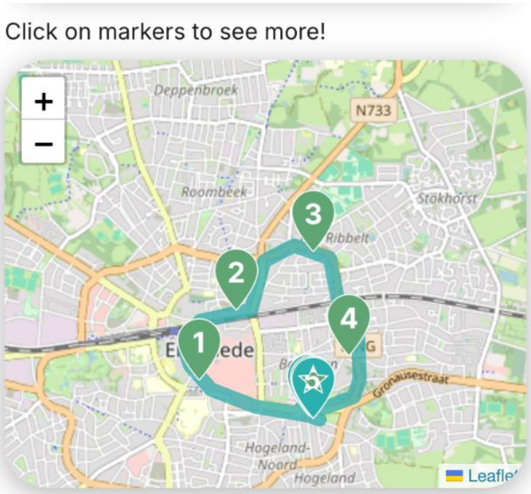
11. If you could change one thing about the experience, what would it be?

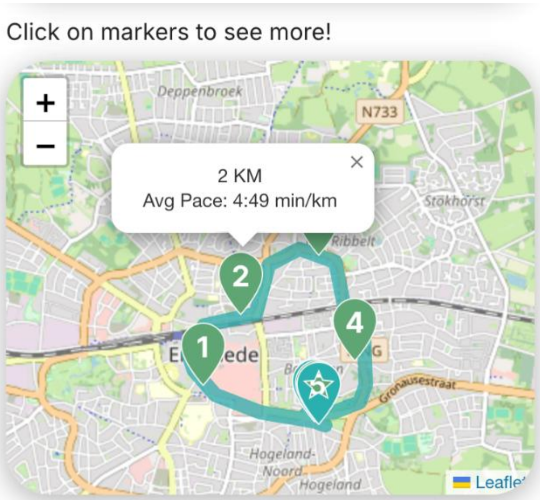
12. Is there anything I didn’t ask that you would like to add?

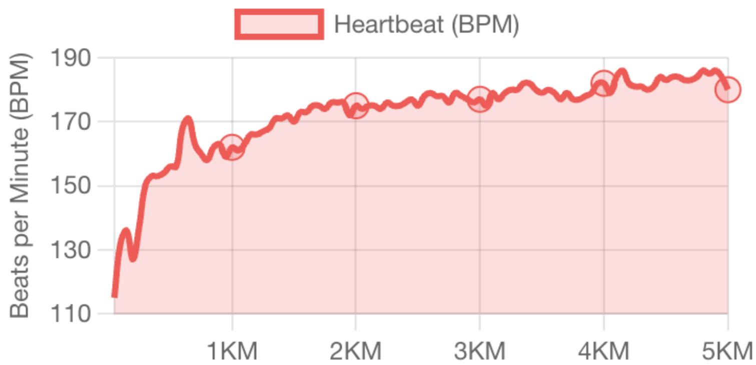
Appendix D

Baseline designs



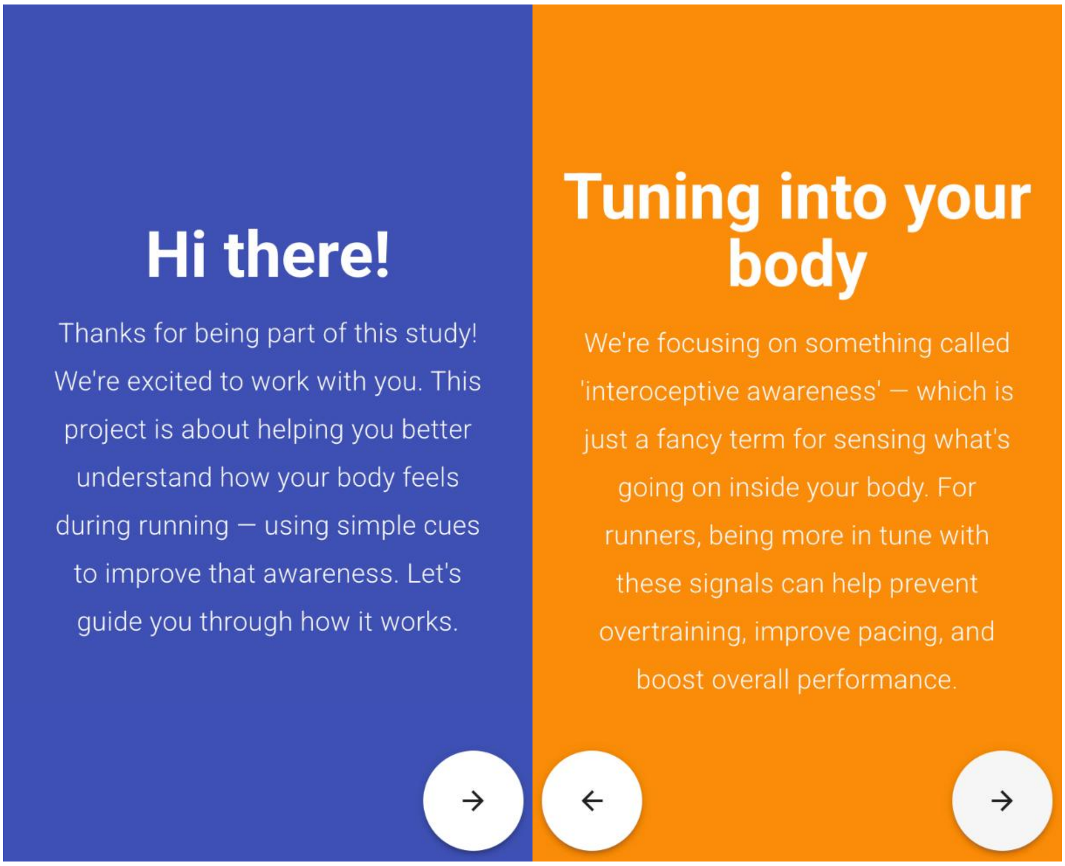


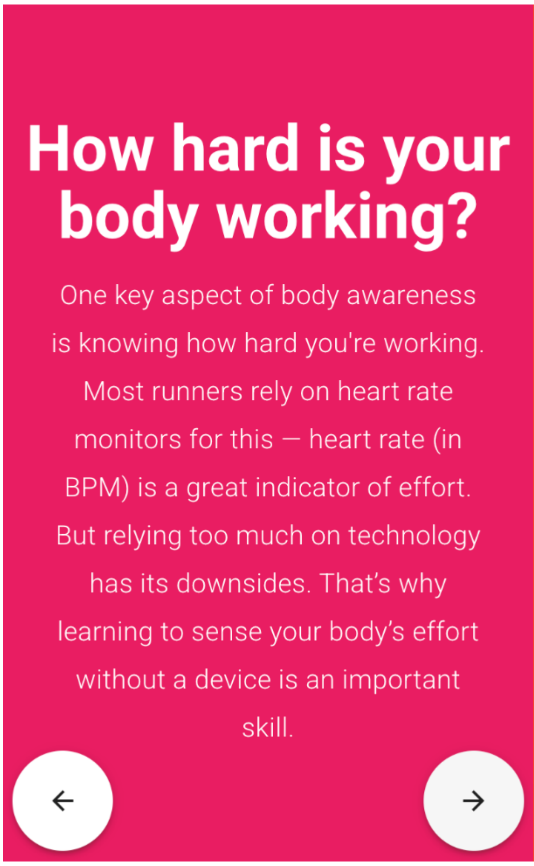


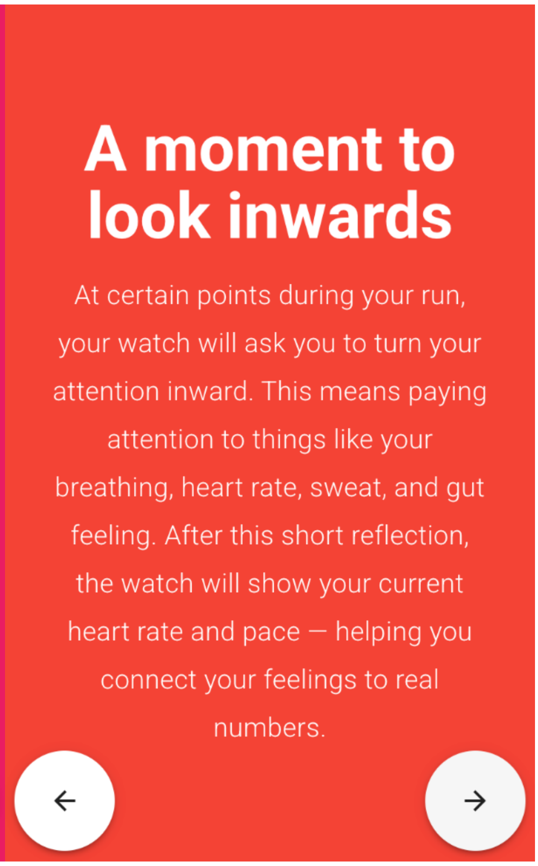


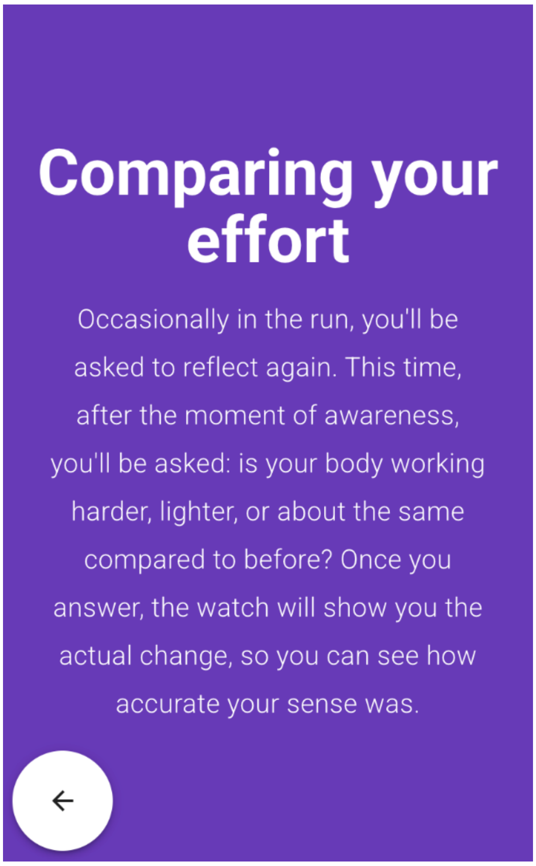
Appendix E

Onboarding




How hard is your body working?
One key aspect of body awareness is knowing how hard you're working. Most runners rely on heart rate monitors for this — heart rate (in BPM) is a great indicator of effort. But relying too much on technology has its downsides. That's why learning to sense your body's effort without a device is an important skill.


A moment to look inwards
At certain points during your run, your watch will ask you to turn your attention inward. This means paying attention to things like your breathing, heart rate, sweat, and gut feeling. After this short reflection, the watch will show your current heart rate and pace — helping you connect your feelings to real numbers.


Comparing your effort
Occasionally in the run, you'll be asked to reflect again. This time, after the moment of awareness, you'll be asked: is your body working harder, lighter, or about the same compared to before? Once you answer, the watch will show you the actual change, so you can see how accurate your sense was.