Studies in Sports Science and Physical Education
ISSN 2959-5274
www.paradigmpress.org/SSSPE
Volume 4 Number 1 2026

The Value and Limitations of Smart-Watch Recovery Scores in Daily Training Decision-Making

Xu Tang[1] & Jiawen Xu[1]

[1] Guangxi Minzu University, Nanning, China
Correspondence: Xu Tang, Guangxi Minzu University, Nanning, China.

doi:10.63593/SSSPE.2026.04.06

Abstract

Smart-watch recovery scores are increasingly used by recreational exercisers to decide whether to train, reduce exercise intensity, or take additional rest. These scores usually combine indicators such as heart rate variability, resting heart rate, sleep, recent training load, and activity history into a single value that appears easy to interpret. Their practical value lies in making recovery-related information more visible and in helping users notice changes that might otherwise be overlooked. However, recovery is not a single physiological condition, and a numerical score cannot fully represent muscular fatigue, subjective tiredness, psychological stress, or the individual response to previous training. Measurement error and differences in proprietary algorithms also limit how precisely these scores can be interpreted. This paper examines the role of smart-watch recovery scores in daily training decision-making, with particular attention to the gap between algorithmic estimates and actual exercise readiness. It argues that recovery scores are most useful when treated as supporting information rather than as direct instructions. Daily training decisions should combine device data with subjective recovery, recent training load, physical discomfort, and the demands of the planned session. Used in this way, wearable technology can improve awareness of recovery without weakening the exerciser's ability to judge and regulate training independently.

Keywords: smart watch, recovery score, training readiness, wearable technology, training decision-making, exercise self-regulation

1. Introduction

Smart watches have gradually moved beyond basic activity tracking. Many devices now provide recovery scores, training readiness estimates, and related feedback to help users decide whether to train, reduce exercise intensity, or rest. For recreational exercisers, these scores are attractive because they simplify several physiological and behavioral indicators into an easily understood result.

The difficulty is that recovery is not a single condition. Sleep, heart rate variability, recent training load, muscular fatigue, soreness, and psychological stress may not change in the same direction. A numerical score can make recovery easier to interpret, but it may also hide important differences between these dimensions. The score itself is also an algorithmic estimate rather than a direct measurement of complete physical recovery.

This creates a practical problem in daily training.

A low score may lead users to reduce training even when they feel well, while a high score may encourage exercise despite fatigue or discomfort. The issue is therefore not whether recovery scores are useful, but how far they should influence training decisions.

This paper examines the value and limitations of smart-watch recovery scores in daily training decision-making. It considers what these scores represent, where their interpretation may become unreliable, and how wearable data can be used without replacing the exerciser's own judgment of recovery and training readiness.

2. What Smart-Watch Recovery Scores Actually Represent

Smart-watch recovery scores are usually presented as simple numerical results, but the information behind them is more complex. Depending on the device, the score may draw on heart rate variability, resting heart rate, sleep duration, recent training load, activity history, and other physiological or behavioral indicators. These data are processed through an algorithm and converted into an estimate of recovery or training readiness. Users normally see only the final number, color, or short recommendation rather than the calculations behind it.

Recovery itself is not directly measured in the same way as running distance or heart rate. A smart watch detects a limited group of signals and uses them as indicators of recovery. Heart rate variability, for example, can reflect changes in autonomic nervous system activity, but it is also affected by sleep, illness, psychological stress, and other conditions. Resting heart rate can provide useful information about changes in physical state, though it cannot identify exercise recovery on its own. Sleep information adds another source of data, yet consumer devices estimate sleep through movement and physiological signals rather than the full assessment used in laboratory settings.

A recovery score is therefore based on several indirect indicators. Combining them can still be useful. One isolated change in resting heart rate may have little meaning, while a pattern involving poor sleep, an unusually high recent training load, and changes in heart-rate-related measures may deserve attention. The score makes such information easier for recreational exercisers to notice without requiring them to interpret every variable separately.

Some dimensions of recovery are much harder for a wrist-worn device to capture. Local muscular fatigue, delayed-onset muscle soreness, joint discomfort, pain, emotional exhaustion, and the subjective feeling of readiness may not be fully represented. A runner may show relatively normal cardiovascular data while the lower limbs remain sore after a demanding session. A person may also receive a low recovery score after one poor night of sleep but still feel capable of completing moderate exercise. The device is describing selected aspects of recovery, not the whole physical state of the user.

Recovery calculations also differ between manufacturers. Devices do not necessarily use the same variables, time periods, or weighting systems. Some place greater emphasis on sleep and heart rate variability, while others incorporate recent training load and additional stress-related information. Similar physiological conditions can consequently produce different readiness estimates on different devices.

The numerical format can make these differences less obvious. A score of 78 or 82 looks precise, although it is still built from sensor measurements, selected variables, and algorithmic assumptions. The number is better treated as a condensed estimate of recovery-related information than as a direct measurement of complete exercise readiness. It can help reveal patterns and changes, but its meaning depends on what the device measures, what it leaves out, and how the available data are interpreted.

3. The Value of Recovery Scores in Daily Training Decisions

Recovery scores can be useful because they bring recovery into everyday training decisions. Recreational exercisers often pay close attention to distance, pace, duration, or calories burned, while recovery receives less attention. A score shown on the watch can make changes in sleep, heart rate, recent training load, or general physical condition more noticeable. Even if the number is not a complete measure of recovery, it can remind the user that the ability to train is affected by what happened before the current session.

This function is especially useful when fatigue develops gradually. A single hard workout is easy to recognize, but several moderately demanding sessions combined with poor sleep may be less obvious. The exerciser may still feel capable of training and continue with the original plan. A declining recovery score can serve as an early signal that recent workload and recovery are no longer well balanced. It does not decide what the athlete should do, but it may encourage a closer look at training intensity, sleep, and physical condition before the next session.

Some smart-watch systems already use recovery information in this way. Huawei, for example, provides estimates of recovery time and recovery percentage after exercise and links different levels of recovery with suggestions about subsequent training intensity. The practical value of this design lies in making recovery easier to understand for users who may not be familiar with physiological indicators. Instead of asking them to interpret several separate measurements, the device gives a simplified indication of whether the body may be ready for another demanding session or whether more recovery could be useful.

Such feedback can also support adjustment rather than complete cancellation of training. Daily training decisions are rarely limited to a choice between hard exercise and total rest. A user with a lower recovery score may reduce running intensity, shorten the session, replace interval training with easy aerobic work, or focus on mobility and technique. Recovery information becomes more useful when it helps modify the planned load rather than producing a rigid “train” or “do not train” response.

The longer-term pattern is often more informative than one isolated score. Recovery data collected over several weeks can help exercisers notice how their bodies respond to different training schedules. Some may find that two consecutive high-intensity sessions are followed by a marked decline in recovery indicators. Others may notice that poor sleep affects the next day’s readiness more strongly than expected. These patterns can gradually improve the way training is arranged. The user is not simply following a daily number but learning how workload and recovery interact over time.

This is one of the stronger arguments for using wearable recovery data. Recreational exercisers usually do not have access to coaches, laboratory testing, or regular physiological assessment. A smart watch cannot replace these methods, but it can provide continuous information at relatively low cost and with little disruption to daily life. The data are available under normal living and training conditions, which makes them useful for observing personal trends that might otherwise go unnoticed.

Recovery scores can also shift attention away from the idea that more training is always better. Fitness technology often encourages users to accumulate steps, kilometers, active minutes, or training load. A recovery indicator introduces a different message: adaptation also depends on rest. This can be particularly valuable for users who tend to increase exercise volume quickly or who interpret every missed session as a loss of progress.

The main value of recovery scores is not that they can tell an exerciser exactly what to do on a given day. Their usefulness is more modest and, in practice, more realistic. They can make recovery visible, draw attention to changes that deserve consideration, and provide another source of information when adjusting daily training. The benefit becomes greater when users look at patterns rather than isolated numbers and relate the score to the actual demands of the planned session. How far that information can be trusted is a separate question, because the apparent simplicity of a recovery score can easily conceal important measurement and interpretive limits.

4. The Limitations of Recovery Scores in Training Guidance

4.1 Measurement and Algorithmic Limitations

The usefulness of a recovery score depends first on the quality of the data used to produce it. Smart watches usually rely on optical sensors, accelerometers, and other wrist-based measurements to estimate heart rate, heart rate variability, sleep, and physical activity. These technologies make continuous monitoring possible outside the laboratory, but convenience does not remove measurement error. Signal quality can change with device position, skin contact, movement, exercise intensity, and individual physiological characteristics. A recovery score built from these signals inherits some of these uncertainties.

Sleep data provide a clear example. Sleep duration and sleep quality are commonly included in recovery or readiness calculations, yet a wrist-worn device does not assess sleep in the same way as polysomnography. A meta-analysis of 24 studies involving 798 participants compared consumer wrist-worn sleep trackers with polysomnography and found significant differences in total sleep time, sleep efficiency, sleep latency, and wake after sleep onset. Another validation study comparing several commercial devices with polysomnography also found systematic bias in sleep estimates, although performance differed across devices and sleep variables. These findings do not mean that wearable sleep data are useless. They show that the values entering a recovery algorithm should not be treated as error-free observations.

Heart-rate-related measures raise a similar issue. Resting heart rate and nocturnal HRV can often be monitored reasonably well under stable conditions, but accuracy does not remain identical across all users and situations. Movement, abnormal heart rhythms, poor sensor contact, and changes in recording conditions can affect the signal. Even when HRV itself is measured accurately, its interpretation is not straightforward. A lower value may reflect training strain, but it can also be influenced by sleep loss, illness, psychological stress, alcohol consumption, or other factors unrelated to the planned workout. The device records a physiological response; the cause of that response still needs interpretation.

Measurement limitations become more important once several variables are combined into one score. Recovery scores belong to a higher level of wearable output. They are not simply sensor readings. A review of consumer wearable technology in cardiovascular and exercise settings noted that metrics such as exercise recovery scores are commonly derived from proprietary algorithms and can differ substantially between manufacturers. The user usually sees the final result but has limited access to the exact weighting of sleep, HRV, training load, stress, or other inputs.

This lack of transparency makes the apparent precision of the score difficult to evaluate. A value of 81 may look more informative than a general judgment such as “moderately recovered,” yet the number does not reveal how much each input contributed to it. It may also be unclear how missing data were handled or how strongly the algorithm responds to an unusual night of sleep. Two watches can collect broadly similar information and still produce different assessments because the underlying models are not the same.

Algorithms are also updated. A user may continue wearing the same device while the way a readiness or recovery metric is calculated changes through software updates. Recent work on consumer wearables has pointed out that proprietary algorithms and limited transparency make independent evaluation difficult, particularly for higher-level outputs such as readiness and recovery scores. This creates a practical problem for longitudinal interpretation. A change in the displayed score may reflect a real change in recovery, but part of the difference can also come from changes in data processing.

Individual variation adds another layer. The same physiological response does not always carry the same meaning for two exercisers. Baseline HRV differs markedly between individuals, and responses to training load are also highly personal. One runner may show a noticeable HRV reduction after a demanding session, while another shows little change. A useful recovery system needs enough personal data to distinguish normal variation from a meaningful departure from the individual baseline. Even then, the result remains probabilistic rather than absolute.

Evidence from training studies also suggests that wearable recovery indicators do not always move in parallel with subjective fatigue. In one study of intensified endurance training, perceived strain and muscle soreness increased during the overload period, while sleep and nightly recovery metrics did not show the same consistent deterioration. This mismatch is important. A recovery algorithm may be functioning exactly as designed and still miss a dimension of fatigue that matters for the next training session.

The measurement problem and the algorithm problem are therefore connected. Sensor data can contain error, and the algorithm then interprets those imperfect signals through a model that is only partly visible to the user. Neither problem makes recovery scores meaningless. It does, however, limit how confidently a single value can be treated as a precise statement about training readiness. The score is better read as an estimate built from selected signals, not as a direct measurement of whether the body has fully recovered.

4.2 The Simplification of Recovery into a Single Score

A recovery score is attractive partly because it reduces complexity. Instead of asking users to interpret sleep, heart rate variability, recent training load, stress, and recovery time separately, the device combines them into one number. Garmin’s Training Readiness, for example, integrates sleep score, recovery time, HRV status, acute load, recent sleep history, and recent stress history into a score from 1 to 100. The result is then translated into simple categories such as poor, low, moderate, high, or prime.

This design has an obvious practical advantage. Most recreational exercisers do not have the knowledge or time to interpret several physiological indicators every morning. A single score lowers the threshold for using recovery information. The problem begins when this convenience is mistaken for physiological completeness. Sleep, cardiovascular recovery, accumulated training load, mental stress, and muscular fatigue do not represent the same process. They may improve or deteriorate at different speeds, yet the final display presents them as one unified state.

The loss of detail becomes important when different recovery dimensions move in opposite directions. A runner may sleep well and show normal HRV but still have substantial lower-limb soreness after a demanding session. Another person may have a poor night of sleep while showing little muscular fatigue and feeling capable of completing moderate exercise. If both situations are reduced to one score, the user sees the final result but not necessarily which part of recovery is driving it.

Weighting also matters. A single score implies that several inputs can be combined into one scale, but users usually do not know how strongly each component influences the final number. Two people may receive the same readiness score for very different reasons. One may have poor sleep but a low recent training load. The other may have good sleep but unusually high accumulated load. Treating the same numerical result as the same physical state overlooks these differences.

The format of the score can also create a sense of certainty that the underlying physiology does not support. A value of 42 appears clearly different from 52, even though the practical meaning of that difference may be uncertain. The category attached to the number can strengthen this effect. When a device labels a score as “low” or “high,” users may interpret the category as a direct judgment of whether training is appropriate, rather than as a simplified summary of selected indicators.

This does not mean that composite scores should be abandoned. Their simplicity is exactly what makes them usable in everyday exercise. The problem lies in treating the simplification as if nothing important had been lost. A recovery score works best as a starting point for interpretation. Users still need to know whether the lower score reflects poor sleep, accumulated workload, stress, or another factor, because different causes may call for different training responses.

The difference matters in daily training decisions. A low score caused mainly by one poor night of sleep may lead to a different adjustment from a low score associated with several days of heavy training and persistent fatigue. The numerical result may be similar, but the appropriate response is not necessarily the same. A single recovery score can summarize information efficiently, but it cannot preserve the full meaning of every component that went into it.

4.3 Conflicts Between Device Scores and Bodily Perception

Recovery scores become most difficult to interpret when the device and the exerciser appear to give different answers. A watch may indicate high recovery or good training readiness while the user still feels heavy, sore, or mentally tired. The opposite can also happen. A low score may appear after poor sleep or an unusual change in HRV even though the person feels capable of completing the planned session. These conflicts are not rare exceptions. They reflect the fact that device-based recovery and bodily perception are built from different kinds of information.

Bodily perception includes signals that are difficult to capture through a wrist-worn sensor. Muscle soreness, local fatigue, joint discomfort, pain, motivation to train, and the general feeling of physical freshness are often experienced directly rather than measured. A runner who completes a demanding downhill session may have considerable quadriceps soreness the next morning. Heart-rate-related measures may already look normal, but the muscles involved in the previous session have not necessarily recovered to the same degree. A recovery score based mainly on cardiovascular and training-load indicators can therefore look more favorable than the athlete feels.

COROS provides a useful example. Its Recovery metric is currently calculated mainly from Base Fitness, Training Load, and the remaining energy estimate following recent exercise. COROS states that sleep, HRV, daily stress, and muscular fatigue are not currently included in this recovery calculation. The company also advises users to pay attention to how they actually feel when subjective fatigue does not match the displayed Recovery value. This example shows why a high recovery percentage should not automatically be interpreted as complete physical recovery. The score describes recovery according to the variables included in the model.

The distinction becomes especially important across different types of exercise. Cardiovascular fatigue and local muscular fatigue do not always recover at the same rate. A long endurance session, heavy resistance training, sprint work, and repeated jumping can produce different patterns of fatigue. A wrist-worn device may capture changes in heart rate or autonomic activity reasonably well while remaining less sensitive to localized muscle damage or joint stress. The same recovery score may therefore have different practical meanings depending on what the athlete did the previous day.

Subjective perception is not perfectly reliable either. Exercisers can underestimate fatigue when they are highly motivated or overestimate it after a poor training experience. Mood, expectations, work pressure, and previous beliefs about recovery can all affect how ready someone feels. The solution is not to replace the watch with subjective feeling. Both sources contain useful information, and both can be incomplete.

Problems arise when the numerical score is automatically given greater authority simply because it looks objective. A device may display a precise percentage, while bodily perception is expressed in vague terms such as “my legs feel heavy” or “I do not feel ready.” The number can seem more scientific, but precision in presentation does not mean that it captures every relevant aspect of recovery. Muscular soreness that is not part of the algorithm does not become less important because it lacks a numerical value.

A conflict between the two sources should instead trigger closer judgment. If the watch reports good recovery but the user has unusual pain, persistent soreness, or marked fatigue, the score should not be used to justify a hard session. If the score is low while the exerciser feels normal, the underlying data can be checked. Poor sleep, an elevated resting heart rate, unusually high recent load, or a change in HRV may explain the result. The decision can then be adjusted according to the type and intensity of the planned exercise rather than following the score mechanically.

The most useful role of recovery technology is not to settle every disagreement between data and perception. It is to make those disagreements visible. A mismatch can reveal that the algorithm has missed something, that the exerciser has overlooked accumulated fatigue, or that one unusual data point is influencing the score more strongly than expected. Learning to interpret these situations is part of exercise self-regulation. Recovery data become more valuable when they sharpen bodily awareness rather than replace it.

5. Using Recovery Scores Without Losing Exercise Self-Regulation

The most practical use of recovery scores is to treat them as one source of information within a broader training judgment. A score can draw attention to changes in sleep, heart rate variability, recent load, or stress, but it cannot decide on its own what type of exercise is appropriate. Daily training still requires the exerciser to consider how the body feels, what was done in previous sessions, and what the planned session is expected to achieve.

Subjective recovery should remain part of this process. Muscle soreness, unusual pain, heavy legs, poor concentration, and a general sense of fatigue can all affect training quality. These signals may not be fully represented in a smart-watch score. If the device reports high readiness while the user has clear physical discomfort, the numerical result should not override that experience. The same principle applies in the opposite direction. A low score does not always require complete rest if the exerciser feels well and the planned session is light. The score needs interpretation rather than automatic obedience.

The type of training also matters. A low recovery score before a high-intensity interval session may justify reducing intensity or moving the session to another day. The same score before an easy walk, light cycling, mobility work, or technical practice may not require the same response. Recovery information is more useful when it helps modify training load than when it produces a simple train-or-rest decision. Volume, intensity, and exercise type can all be adjusted separately.

Looking at trends can also reduce overreaction to single-day changes. One unusual night of sleep may lower a recovery score without indicating a meaningful decline in physical condition. A pattern that continues for several days is more difficult to dismiss, especially when it appears together with poor training performance, persistent fatigue, or changes in resting heart rate. Users who learn to read these patterns are more likely to use wearable data as context rather than as a daily command.

The relationship between data and personal experience can become more useful over time. An exerciser may notice that a certain level of training load usually produces two days of lower recovery scores, or that poor sleep has a stronger effect on readiness than expected. Another person may find that device scores recover quickly after strength training even though muscular soreness remains. These personal patterns help explain what the score means for that individual. The number becomes more informative when it is compared with repeated experience rather than interpreted in isolation.

Exercise self-regulation also involves knowing when wearable feedback should carry less weight. Pain, symptoms of illness, unusual cardiovascular responses, or prolonged fatigue require more caution than an ordinary change in recovery score. A high readiness value cannot make these signs irrelevant. Recreational exercisers are especially vulnerable to this problem because they may not have a coach or medical professional available to interpret conflicting information.

There is also a risk that constant monitoring changes the way people understand exercise. If every session begins with checking whether the watch gives permission to train, users may gradually lose confidence in their own judgment. They may also become anxious when the score is lower than expected, even when there is no clear physical problem. Wearable technology then moves from supporting exercise management to controlling it.

A better outcome is that repeated use of recovery data improves the exerciser's ability to recognize personal patterns. The watch can help users connect heavy training with later fatigue, identify the effect of poor sleep, and notice when recovery takes longer than usual. Over time, this should make training judgment more informed, not more dependent on the device.

The strongest use of recovery scores is therefore not to follow them exactly. It is to place them beside subjective recovery, recent training load, current physical condition, and the demands of the next session. When these sources point in the same direction, the decision may be relatively straightforward. When they conflict, the disagreement itself becomes useful information. Exercise self-regulation is preserved when the device contributes to judgment without becoming the final authority over it.

6. Conclusion

Smart-watch recovery scores can make recovery-related information easier to notice and use in daily training. They bring together data such as sleep, heart rate variability, recent training load, and activity history, giving recreational exercisers a simple reference when deciding whether to train, reduce intensity, or recover.

Their usefulness, however, has clear limits. Recovery is multidimensional, while the score is built from selected variables and proprietary algorithms. Muscular fatigue, pain, psychological state, and subjective readiness may not be fully represented. A precise-looking number can therefore simplify a condition that is far less precise in practice.

Recovery scores are most useful when they support rather than replace training judgment. Device data should be considered together with recent workload, physical discomfort, subjective fatigue, and the demands of the planned session. Trends across several days are usually more meaningful than one isolated result.

The aim of wearable monitoring should not be to make exercisers dependent on a daily score. Its greater value lies in helping them understand their own recovery patterns and make better training decisions with less uncertainty. When the device becomes part of that learning process rather than the final authority, recovery technology can support more informed and sustainable exercise self-regulation.

References

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., Gabbett, T. J., Coutts, A. J., Burgess, D. J., Gregson, W., & Cable, N. T. (2017). Monitoring athlete training loads: Consensus statement. *International Journal of Sports Physiology and Performance, 12*(Suppl. 2), S2-161–S2-170.

Cardinale, M., & Varley, M. C. (2017). Wearable training-monitoring technology: Applications, challenges, and opportunities. *International Journal of Sports Physiology and Performance, 12*(Suppl. 2), S2-55–S2-62.

Düking, P., Fuss, F. K., Holmberg, H.-C., & Sperlich, B. (2018). Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. *JMIR mHealth and uHealth, 6*(4), e102.

Halson, S. L. (2014). Monitoring training load to understand fatigue in athletes. *Sports Medicine, 44*(Suppl. 2), S139–S147.

Ibrahim, A. H., Beaumont, C. T., & Strohacker, K. (2024). Exploring regular exercisers' experiences with readiness/recovery scores produced by wearable devices: A descriptive qualitative study. *Applied Psychophysiology and Biofeedback, 49*(3), 395–405.

Lee, Y. J., Lee, J. Y., Cho, J. H., Kang, Y. J., & Choi, J. H. (2025). Performance of consumer wrist-worn sleep tracking devices compared to polysomnography: A meta-analysis. *Journal of Clinical Sleep Medicine, 21*(3), 573–582.

Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M. (2013). Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. *Sports Medicine, 43*(9), 773–781.

Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: A systematic review. *British Journal of Sports Medicine, 50*(5), 281–291.