Journal of Physical Education and Sport ® (JPES), Vol. 25 (issue 10), Art 221, pp. 1985 - 1995, October 2025
online ISSN: 2247 - 806X; p-ISSN: 2247 – 8051; ISSN - L = 2247 - 8051 © JPES

Original Article

Validation of a Trail Segmentation and Analysis Tool (TSAT) using Strava segmentation

DE COCK FLORENCE[1], CAILTEUX EMILIE[2], PARTOUNE MARIE[3], JIDOVTSEFF BORIS[4]
[1,2,3,]Department of Sport and Rehabilitation Sciences, Research Unit EVA-REVA, Faculty of medicine, University of Liège, BELGIUM
[4]Department of Sport and Rehabilitation Sciences, Research Unit for a Life-Course Perspective on Health and Education - RUCHE, Faculty of medicine, University of Liège, BELGIUM

Published online: October 31, 2025
Accepted for publication: October 15, 2025
DOI:10.7752/jpes.2025.10221

Abstract
Purpose: Currently, there is no analytical tool specifically designed to manage performance and effort in trail running. To address this gap, we developed the Trail Segmentation and Analysis Tool (TSAT), which uses GPX data and Strava-based segmentation. The TSAT consists of two components: the Trail Identity Card and the Trail Effort Analysis. This study aimed to validate the tool's content through expert evaluation. **Method:** Ten experts participated in the validation process of the TSAT through semi-structured interviews. They responded to closed-ended questions (analyzed using the Content Validity Index) and provided qualitative feedback through open-ended responses. **Results:** Experts unanimously found the TSAT to be reliable, relevant and useful. The Content Validity Index confirmed strong support for most components, although some thresholds were questioned due to limited empirical evidence. Qualitative feedback led to refinements, such as clearer definitions of trail technicity and broader segment categories. **Conclusions:** This research enabled the validation of the overall content of the TSAT. This tool offers practical applications for athletes, coaches, and race organizers by enabling detailed performance analysis. While the tool shows strong initial validation, a few minor modifications have been made, and further studies are needed to refine thresholds.

Keywords: **Running, Effort management, Performance, Experts**

Introduction

The popularity of endurance sports has grown significantly in recent years, with trail running emerging as one of the fastest-growing disciplines worldwide (Hoffman, 2010; Scheer, 2019; Stöhr et al., 2021). Unlike road running, trail running takes place predominantly in natural environments, featuring technical terrain, substantial elevation changes, natural obstacles (e.g., rocks, roots, mud), and variable weather conditions. Races vary in distance from a few kilometers to ultra-trails exceeding 150km, with difficulty often classified using the kilometer-effort unit, which accounts for both distance and elevation gain (*ITRA Discover Trail Running*, s. d.).
Traditional performance prediction models for track and road running are not adapted for trail running due to its unique demands (Ehrström et al., 2018; Koerner, 2019). In addition to distance, trail running performance depends on elevation gain, terrain technicity, and the need for advanced motor control (Chase & Hobbs, 2023; Santuz et al., 2018). The unpredictable nature of trails requires runners to constantly adapt their speed, stride, and gait; sometimes resorting to walking on steep or technical sections (Ehrström, 2020; Zimmermann et al., 2022). Furthermore, environmental factors such as weather and unstable ground further complicate performance analysis, making conventional models inadequate.To comprehensively study trail running performance, detailed data on race characteristics and effort management must be collected. Modern GPS technology also allows for precise tracking of runners' pacing strategies across the race, facilitating in-depth performance evaluations (Czegledi et al., 2023). Strava is an application that uses runners' GPX file records to provide an overview of their effort. In particular, the application has developed a very popular function that displays a runner's time over a defined segment of the race. This segment-based approach is interesting because it allows a race to be broken down into segments with similar characteristics (such as climbs, descents or flat sections), offering the possibility of analyzing effort throughout the race, taking account of its characteristics.

The Strava application offers an original and relevant solution by enabling the creation of segmented race analysis. Recent research has demonstrated Strava's high accuracy, with measurement errors below 2% for segments at least 200 meters long, particularly in open terrain (De Cock et al., 2023).

Despite the growing popularity of trail running and increasing scientific interest, no dedicated tool currently exists to systematically analyze trail characteristics and effort management. To address this gap, this study aims to develop and validate an original analytical tool based on GPX race traces and Strava segmentation: the **Trail Segmentation and Analysis Tool (TSAT)**. We conducted expert panel validation to ensure the tool's relevance and clarity, laying the groundwork for future research on trail running performance determinants and individualized pacing strategies.

Corresponding Author: DE COCK FLORENCE, E-mail: fdecock@uliege.be

Methods

This study was carried out in three successive stages: (1) initial creation of the tool; (2) content validation by a panel of experts; (3) revision and finalization of the tool based on experts' feedback. This progressive methodological approach ensured the relevance, clarity and reliability of the tool developed.

Development and presentation of the tool

The TSAT was developed at the University of Liège to analyze short-trail efforts while considering the specific nature of the race. The TSAT combines pre-race course data with participants' race-day performance metrics. The system begins with a GPX recording of the race route, which is segmented in Strava to identify key sections (climbs, descents, and flats) based on slope gradients and minimum distance criteria. During the actual race, participants' GPS devices record their performance data, which is automatically synchronized with Strava. After the event, each runner's segment times are extracted from Strava and compiled into a customized Excel-based analysis tool. This tool incorporates automated functions to process the data, enabling semi-automatic calculation of performance metrics. The tool is based on two complementary components: (A) a **trail identity card** (TIC); (B) a **trail effort analysis** (TEA) of the runner during this race.

A. Trail Identity Card (TIC)

The TIC (Figure I) provides a comprehensive profile of the trail race, derived from pre-race reconnaissance and GPX data; and includes **basic metrics** such as:

1. total distance of the trail (km)
2. average altitude of the trail (m); positive/negative elevation gain (D+/D−, in meters)
3. kilometer-effort (KmE), calculated as:

$$KmE = Distance\ (km) + \left(\frac{D^{+}(m)}{100}\right)$$

4. ITRA (International Trail Running Association) category: classified by KmE (from XXS to XXL)

The TIC also provides a **Technicity Score** of the race. To do this, the local technicity of each segment is recorded during the race day itself. This technicity is obtained with an original technicity scale which we have developed (Appendix I). This scale, from 0 to 10, is irrespective of terrain. It is determined by the accumulation of factors over the race that impede the runner's running speed and normal progress. These may include uneven ground (stones, rocks, ditches, etc.), difficult supports and grips (mud, sand, grass, etc.), slippery terrain (wet ground, snow, etc.), or obstacles to overcome (tree trunks, stream, etc.). A highly technical course will require great adaptation on the part of the runner (modified foot position, adaptation of support, change of pace, walking, climbing, special vigilance). The technicity data for each segment is then encoded and a weighted average score is obtained, characterizing the overall technicity nature of the race.

The TIC also offers a **view of the route profile** (obtained by the GPX trace of the route) and details on course **segmentation** from the Strava application. Four different types of segments were determined, mainly based on the profile of the course:

1 **Uphill**: slope over 3%, positive, 2. **Downhill**: slope greater than 3%, negative, 3. **Flat**: slope under 3% (positive and negative), 4. **Undulating**: alternating short negative and positive slope over 3%

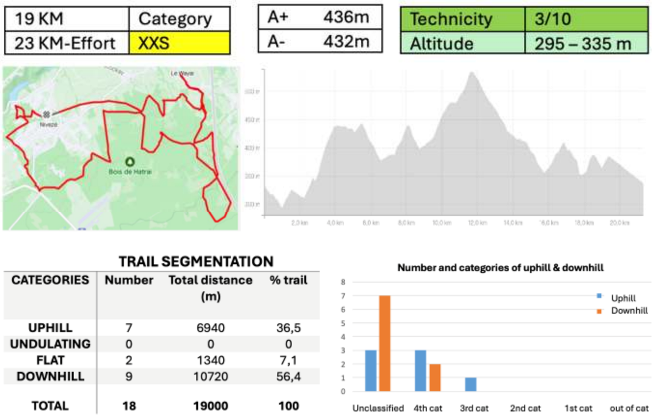
Preliminary work has established that for a quality analysis, segments should be at least 400m long, except for very steep segments (slope >10%) for which a minimum distance of 200m is required.

Finally, the TIC gives information on the number and categories of uphill (and downhill) segments according to the model used in cycling for mountain passes:

Score = length x (gradient)²

This calculation procedure is widely recognized and used in international cycling competitions (Rodriguez-Marroyo et al., 2003). There are six different categories: "unclassified" (<35pts); "category 4" (35-79pts); "category 3" (80-179pts); "category 2" (180-249pts); "category 1" (250-599pts) and "out of category" (>600pts). The categorization aims to provide a complete and clear presentation of the level of difficulty of the uphill and downhill encountered during the race.

Figure I. Example of a trail identity card (TIC)



19 KM	Category
23 KM-Effort	XXS

A+	436m
A-	432m

Technicity	3/10
Altitude	295 – 335 m

TRAIL SEGMENTATION

CATEGORIES	Number	Total distance (m)	% trail
UPHILL	7	6940	36,5
UNDULATING	0	0	0
FLAT	2	1340	7,1
DOWNHILL	9	10720	56,4
TOTAL	18	19000	100

B. Trail effort analysis of the runner (TEA)

The TEA (Figure II) evaluates runner performance relative to course features using the original concepts of **Expected Speed (ES) and Relative Speed Deviation (%RSD). ES** is a normalized benchmark representing the speed a trail runner should maintain on a specific segment if their effort is uniformly distributed across the entire race. It adjusts for individual performance relative to the group average, enabling segment-by-segment comparison of pacing strategies. ES accounts for the runner's overall performance level (e.g., a faster runner's ES will be higher than a slower runner's for the same segment) and is derived from the group's average speed on each segment, scaled to the runner's total race time. To determine ES, we first need to compute the runner's **Expected Segment Time** for each segment:

$$\textit{Expected Segment Time} = \left(\frac{\text{Runner's Total Time}}{\text{Group Avg. Total Time}}\right) \times \textit{Group Avg. Segment Time}$$

Runner's Total Time is the sum of the times (in seconds) measured for each segment of the trail while **Group Avg. Total Time** is the sum of the average times measured for each segment for the group of runners participating in the study.

ES (in m/s) can then be computed as follow:

$$ES = \frac{\textit{Segment Distance}}{\textit{Expected Segment Time}}$$

Then we calculated the **Relative Speed Deviation (%RSD)** for each segment as:

$$\%RSD = \left(\frac{\textit{Measured Speed (MS)}}{\textit{Expected Speed (ES)}} - 1\right) \times 100$$

The %RSD metric measures the relative deviation (in %) of a runner's measured speed (MS) in comparison to their ES for each segment. A %RSD of 0% means the runner's speed in the segment corresponds to the ES; a positive value indicates a faster pace than expected and a negative value reflects a MS slower than the ES. This approach transcends conventional pace analysis by contextualizing speed within individualized expectations. By quantifying deviations from a runner's predicted segment performance, it isolates terrain-specific adaptations, pacing errors, and physiological limits, offering a granular toolkit for optimizing trail running strategies.

Figure II. Example of an individual trail effort analysis of the runner (TEA)

Participant	**Part 1**

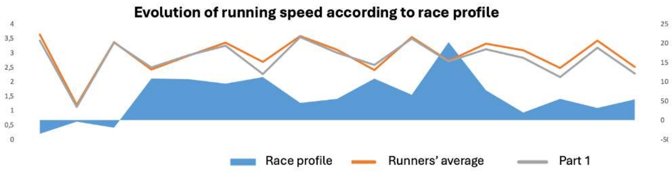
19 KM	Category
23 KM-Effort	XXS

A+	436m
A-	432m

Technicity	3/10
Altitude	295 – 335 m

Chronometer	**1h50min53sec**
Average speed	**10,3 km/h**
Average speed of all runners	10,6 km/h

Weather conditions				
Covering	**T°**	**Humidity**	**Wind speed**	**Specific conditions**
Cloudy	16°C	98%	19,4km/h	/

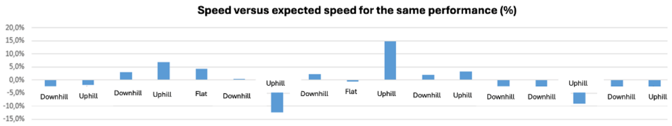


Analysis based on race profile		
Segments	**Speed**	**% Expected Speed**
Uphill	8,7km/h	2,3%
Undulating	/	/
Flat	10,6km/h	1,6%
Downhill	11,6km/h	-2,2%

Strength	Uphill
Weakness	Downhill
Type of runner	Balanced

Overall effort management					
	1st quartile	2nd & 3rd quartile		4th quartile	Variability of Expected Speed
% Expected Speed	3,0% Start	2,5% Middle	0,4% Middle	-5,7% Finish	4,0% All the race
Conclusion	Standard	Standard	Standard	Slow	Normal variations

Differences between the first and second halves of the race	
-10,8%	Positive Split



The division of the trail into segments and the categorization of these segments proposed by the TSAT makes it possible to carry out two types of analysis of effort management in trail running: 1) a specific analysis by segment category and 2) an analysis of pace management throughout the race.

1. Analysis by segment category

For the same runner, the average %RSD achieved over the same categories of segments is grouped, giving an indication of the runner's strengths and weaknesses per category of segment (extreme values) analyzed. The difference between uphill and downhill is used to determine the type of runner, and the 5% threshold was used:

- **Climber runner**: %RSD uphill - %RSD downhill ≥ 5%, **Descender runner**: %RSD uphill - %RSD downhill ≤ -5%, **Balanced runner**: 5% > %RSD uphill - %RSD downhill > -5%

2. Analysis of pace management throughout the race

Three analyses offered by the TSAT enable an analysis of running rhythm management throughout an effort. A first analysis divides the race into **four quarters** (25% of the trail distance) to analyze how the race pace changes from one quartile to another. The first quarter (0-25%) is used to analyze the start of the race, the second and third quarters (25%-50%-75%) are used to analyze the middle of the race and the fourth quarter (75-100%) is used to analyze the end of the race. A 3%RSD threshold was set to consider any part (start, middle and end) as fast (if > 3%) or slow (if < -3%). The second analysis compares the pace between the first and second halves of the race, seeking to verify a positive or negative split strategy. A **positive split** strategy is considered when the second half is at least 5% slower than the first half, and a **negative split** strategy is considered when the second half is at least 5% faster than the first half. Within the literature, there is no universally accepted threshold for determining the significance of this difference, as it can vary according to race distance and specific conditions. However, a difference of a few percent (around 2-3%) between the two halves is often considered significant. We therefore set the threshold at 5% to be sure of observing a real difference in pace between the two halves of the race. The third analysis studies the **variation in pace** segment by segment throughout the race. The standard deviation (SD) of each segment %RSD was used to consider the level of pace variation. The following thresholds were arbitrarily set for interpretation as follows: 1) small variations (SD < 3%); 2) normal variations (3 < SD < 5%); high variations (SD > 5%).

Content validation of the TSAT by a panel of experts

Following the initial development of the TSAT, a content validation process with independent expert reviewers was conducted to verify its scientific relevance and practical applicability. This step is critical to ensure that the tool's components accurately represent key aspects of trail running performance analysis.

Recruitment of experts

The experts were tasked with carrying out a critical analysis of the tool as a whole, as well as for each individual component. They had to have experience in trail running and/or scientific research related to the research question. They had to be fluent in French, which was the language of the researchers. To recruit these experts, an information letter presenting the research and its objectives was sent to a series of identified individuals who met the required expertise criteria. Recruitment was carried out on a voluntary basis, resulting in the participation of 10 experts with varied and complementary profiles, presented in Table I.

Table I. Experts' characteristics

	Educational background	Formation	Scientific expertise	Expertise questionnaire	Running and trail expertise	Running and trail amateur
Exp1	Master	Sports training	Yes	No	Yes	Yes
Exp2	PhD	Physical education + physiotherapy	Yes	Yes	No	Yes
Exp3	PhD	Physical education	Yes	Yes	No	Yes
Exp4	Bachelier	Educational sciences	No	No	Yes	Yes
Exp5	PhD	Physical education	Yes	Yes	Yes	Yes
Exp6	PhD	Physical education	Yes	Yes	Yes	Yes
Exp7	Bachelier	Mechanical sciences	No	No	Yes	Yes
Exp8	Master	Physical education + physiotherapy	Yes	Yes	Yes	Yes
Exp9	PhD	Physical education	Yes	No	Yes	Yes
Exp10	Master	Physical education + physiotherapy	No	No	Yes	Yes

Interviews with the experts

To gather the experts' opinions, we chose a semi-structured interview approach, which allowed for a mixed analytical approach based on quantitative and qualitative data. To access the views of international experts, we decided to carry out the interviews by videoconference using the Teams platform (Microsoft). This approach enabled the tool to be presented in detail using a PowerPoint presentation and screen sharing. Teams was also used to automatically transcribe the recorded interviews (with the agreement of the experts). A researcher systematically reread the transcripts to correct any errors in the Teams automated system. The interviews lasted approximately an hour and began with a general presentation of the tool and its objectives.

To gather the experts' opinions on the tool, we produced a presentation that illustrated the tool as a whole, as well as its various components. For each important element, we followed the same logic, which consisted of first asking the expert about their level of agreement (closed question) and then inviting them to comment on their answer in an open manner, especially when they did not completely agree with the proposal made. The interview was structured similarly way to the presentation of the tool used in the method above, with the first part of the interview dealing with the trail identity card, and the second with the analysis of the effort. The various questions asked are shown in Tables II and III.

Data processing

Following the interviews, the experts' answers to closed questions were encoded into an Excel file. Based on a Likert scale, a score was assigned for each question according to the level of agreement (strongly disagree = 0pts; somewhat disagree = 1pts; somewhat agree = 2pts and strongly agree = 3pts). If the answer was “no opinion”, the table cell was left blank. The Content Validity Index (CVI) was used to assess the degree of relevance for each item in the tool, as well as for the more global questions on its content and form. This method is appropriate for the content validation process of the tool (Yusoff, 2019). An item is deemed relevant when the CVI reaches a score greater than or equal to 0.8. Validation was not limited to this score; it also incorporates expert comments that were considered useful for improving the tool. Secondly, a qualitative approach was conducted for the open-ended questions. The researchers undertook a careful analysis of the experts’ comments to identify and classify the key emerging ideas according to their recurrence and relevance. The researchers considered elements “very relevant” when they had a high potential for improving the tool; “moderately relevant” when they required further reflection and the potential for improving the tool was more uncertain; and as “irrelevant” when there was no interest in further analysis to improve the tool.

Revision of the tool based on the experts’ feedback

Both quantitative (CVI) and qualitative (comments classified as “very relevant” and “moderately relevant”) information were used to identify problems with the tool and areas for improvement. Questions with a CVI <0.8 received greater attention. The changes suggested by the results were discussed within the research team to reach a consensus on the changes to be made to the tool.

Results

Validation of the Trail Identity Card (TIC)

The TIC, which presents key characteristics of the race, was largely validated by the panel of experts (Table II). The proposed structure (including total distance, elevation gain, KmE, and segmentation by terrain type) was judged relevant, with a perfect CVI score of 1.0. Experts appreciated the clarity of the presentation and the integration of meaningful data. *“All in all, I think it's a nice visual to place at race registration or on Instagram/Facebook”* (E8). One minor improvement suggested was the addition of the number of aid stations, which is important for interpreting effort variation across segments (E8). *“Don't forget that in trails, there are aid stations. The runner stops and eats/drinks, so it's part of a segment. This will have an influence on the average speed in relation to the expected speed on the segment.”* (E8). This modification was subsequently integrated into the improved version of the tool.

Table II. Main results (CVI and qualitative analysis) obtained for each question on the Trail Identity Card (TIC) and suggestions for improvement

Questions	CVI	Comments mentioned by experts (Qualitative analysis)	Improvements proposed by the research team
Relevance of the proposed presentation of TIC	1	Relevant and clear Add the number of aid stations (E8)	Adding the number of aid stations
Knowledge of a method that links total distance and positive/negative altitude	/	No other method than km-effort You use the most frequently	
Relevance of the proposed assessment of technicity	0,9	Relevant and easy to understand No other technicity scale known Reduce subjectivity: Excluding weather conditions (E7 and 8) More specific indicators for each score (E3) Standardization of terms (E3) Adding commitment and vigilance (E5)	All points have been retained: technicity under dry conditions using clearly defined indicators across a 0–10 scale
Relevance of the 4-category of segments	1	Appropriate Splitting the uphill and downhills into two sub-categories (easy and steep) (E6 and 9)	Six distinct segment categories are established including “steep uphill” and “steep downhill”
Relevance of the proposed 3% slope thresholds	**0,6**	3% with a limit of 10 or 12% (E6 and 9)	3% and 10% slope thresholds are used
Relevance of the proposed minimum distance of 400m and its exception	0,9	Relevant	
Relevance of the proposed categorisation of uphill/downhills and the calculation	0,9	Relevant but more appropriate in the mountains than in Belgium (E4, 6, 8 and 10)	

The technicity scale, an original feature of the TIC, also received strong support (CVI = 0.9). Experts agreed on its relevance and clarity but proposed several refinements to reduce subjectivity. Several experts (E5, 6, 7, 8 and 9) confirm that the weather on race day can vary this technicity. *“Depending on the weather, it can* *also change, and you have to take that into consideration. It can have an impact on the level of technicity: roots when it's dry or when it's wet, it's not necessarily the same thing"* (E6). Specifically, some experts (E7 and 8) recommended removing references to weather conditions (e.g., mud, snow) to focus the score on terrain-related obstacles only. Suggestions also included standardizing terminology and defining levels more precisely. *"For elements from 1 to 6, we talk about obstacles that reduce very slightly or moderately, and then afterwards, these obstacles are no longer present. On the other hand, specific indicators should be added to each score to clearly identify each level"* (E3). Another expert (E5) also suggests incorporating aspects such as runner commitment and injury risk. *"From a high level of technicity, there's clearly a need for control, which brings with it the risk of falls, sprains or other hazards. I think there could be an indicator of the level of commitment or precaution required"* (E5). These comments led to a revised version of the scale (Appendix II), which now assesses technicity under dry conditions using clearly defined indicators across a 0–10 scale.

Segment categorization was unanimously considered appropriate (CVI = 1.0), especially the distinction between flat, uphill, downhill, and undulating sections. However, the 3% slope threshold used to define segment types was questioned (CVI = 0.6). Several experts (E6 and 9) felt it lacked nuance, proposing an additional threshold to distinguish between moderate and steep slopes (e.g., above 10%). *"It might be a good idea to separate an easy uphill which could be between 3% and 10%, and then a steeper uphill above 10%"* (E6). This suggestion was considered relevant, and the improved version of the tool now integrates six segment types, including "steep uphill" and "steep downhill," reflecting these suggested distinctions. While the categorization of uphill/downhill segments using a cycling-inspired scoring method (based on gradient and length) was also validated (CVI = 0.9), many experts (E4, 6, 8 and 10) noted that its application may be more relevant to mountainous trails than to flatter terrains such as those in Belgium. *"I think this method makes sense. I tend to agree, but you have to realize that here in Belgium, you won't find all these categories"* (E10).

Validation of the Trail Effort Analysis (TEA)

The TEA, which aims to evaluate runners' pacing strategies in relation to trail characteristics, was also positively received by the expert panel (Table III). The proposed segment-by-segment analytical approach reached a CVI of 0.9, with experts highlighting its potential to identify specific pacing patterns. The concept of ES, serving as a personalized benchmark derived from the group average and scaled to each runner's total time, was also validated (CVI = 0.9). Experts (E4, 6, 7, 8, 9 and 10) considered it strong foundation for individualized analysis. *"It's good and complete. I've never seen so much analysis"* (E7); *"It's a nice report on a race; I find it very interesting"* (E10). One expert (E6) suggested refining the benchmark further by comparing each runner not to the entire group, but to a sub-group of peers with similar performance levels. "*What would be interesting is not so much to look at the average group, but rather at the runners who are at his level. For example, if we have five runners faster than him, five slower than him, and take the average rather than the whole group. That way, he can see where his strengths and weaknesses lie in relation to people with the same speed as him"* (E6).

Table III. Main results (CVI and qualitative analysis) obtained for each question on the Trail Effort Analysis (TEA) and suggestions for improvement

Questions	CVI	Comments mentioned by experts (Qualitative analysis)	Improvements proposed by the research team
Relevance of the proposed segment-by-segment analytical approach	0,9	Relevant Add the presence of aid stations (E8)	Adding the number of aid stations
Relevance of the concept of expected speed	0,9	Strong foundation for individualized analysis (E4, 6, 7, 8, 9 and 10)	
Relevance of the approach to highlight segments that were run faster or slower	1	To be compared with a sub-group of peers (5 runners) with similar performance levels (E6)	
Relevance of the classification of runners into 3 categories	**0,6**	Add a 4th type of runner, corresponding to someone who performs very well on the flat (E9 and 10)	
Relevance of the proposed 5% and -5% thresholds	**0,5**	4 experts have no opinion (E3, 4, 6 and 10) (lack of empirical basis and calling for further data-driven validation) Raise thresholds (E9)	
Relevance of the proposed breakdown of the race (beginning, middle and end)	0,9	Adapting this breakdown based on race length (short vs long trails) (E1 and 3) Rename parts as quarters (E5)	
Relevance of the proposed 3% threshold	**0,7**	3 experts have no opinion (E1, 3 and 8) (lack of empirical basis and calling for further data-driven validation)	
Relevance of the analytical approach to the positive/negative split	0,9	Relevant Perhaps to be reviewed for long trails (E3)	
Relevance of the proposed 5% threshold	**0,4**	5 experts have no opinion (E1, 2, 3, 4 and 10) (lack of empirical basis and calling for further data-driven validation)	
Relevance of the concept of speed variation	0,8	2 experts have no opinion (E3 and 8) (lack of empirical basis and calling for further data-driven validation)	
Relevance of the proposed thresholds for speed variation	**0,6**	4 experts have no opinion (E3, 4, 6 and 8) (lack of empirical basis and calling for further data-driven validation)	
Relevance and opinion about the whole concept	1	High level of interest and valuable tool (E4, 8 and 9) Total dependence on the Strava platform (E5 and 10) Potential developments such as gender-specific analyses? (E8) An app-based version for automated use? (E10)	

The method for identifying over- or under-performance on each segment using %RSD was fully validated (CVI = 1.0). However, the proposal to classify runners into three categories (climber, descender, balanced) based on differences in uphill and downhill %RSD received a lower CVI of 0.6. Experts (E9 and 10) proposed adding a fourth category to account for runners who perform particularly well on flat terrain. Additionally, the 5% thresholds used to define runner types were considered arbitrary and insufficiently supported by evidence (CVI = 0.5). For one expert (E9), this threshold could be increased, whereas several experts (E3, 4, 6 and 10) refrained from giving an opinion on these thresholds, citing a lack of empirical basis and calling for further data-driven validation.

The section considering to pacing strategy (based on segment analysis over race quarters, positive/negative split assessment, and variability of speed) was generally validated. The division of the race into four quarters for assessing pacing patterns received a CVI of 0.9. Nonetheless, some experts (E1 and 3) recommended adapting this breakdown based on race length. For example, dividing short trail races into thirds might be more appropriate, while longer races justify the quarter-based model. *"For very long trails, it could be restrictive to place yourself at the beginning, middle and end, because you're going to be in periods that can last more than 10h. In those 10hours, there are going to be moments of good or bad race management*" (E3). One expert (E5) also advised renaming "beginning", "middle" and "end" of race to "first quarter," "middle quarters," and "final quarter" to avoid misinterpretation.

The analysis of positive vs. negative splits (CVI = 0.9) was deemed relevant for assessing pacing strategies, although one expert (E3) questioned its validity in long-duration trail running, where fatigue and environmental factors may play a greater role. "*When running on the flat or over reasonable distances, effort management means that you can be in a positive or negative split voluntarily. I'm questioning this concept a little in long-distance trail running. Is it going to be relevant enough for trail runners?"* (E3). The threshold used to define splits (5%) received the lowest CVI (0.4), with many experts (E1, 2, 3, 4 and 10) either disagreeing with its magnitude or declining to take a position due to insufficient empirical evidence. Similarly, while the concept of speed variation throughout the race (based on standard deviation of %RSD) was validated (CVI = 0.8), the interpretation thresholds (<3%, 3–5%, >5%) scored a modest CVI of 0.6, again due to hesitancy in endorsing fixed cutoffs.

Overall validation of the TSAT

The overall structure and purpose of the TSAT were strongly endorsed by all the experts. The final question assessing the tool's global relevance received a CVI of 1.0, and the qualitative feedback confirmed a high level of interest and perceived utility. Experts (E4, 8 and 9) described the TSAT as a valuable tool for analyzing performance and effort management in trail running, with potential applications in coaching, race planning, and self-assessment. "*If I look at it from the athlete's point of view, I'd be interested in this kind of data to improve on weaknesses*" (E4); *"It's a slightly innovative concept that could perhaps be used for more detailed training planning"* (E8); *"I think it's interesting! It could be used by runners, to offer them a comparison with their previous races"* (E9). Several experts also raised questions about the future of the tool, suggesting potential developments such as gender-specific analyses (E8), integration of environmental factors, or even an app-based version for automated use (E10). Others (E5 and 10) noted that while the tool currently depends on Strava for data extraction, future iterations could explore more autonomous systems. *"Will it be available? Do you envisage an application or structure that will enable the automation of the tool?"* (E10).

Discussion

The present study aimed to validate the content of the TSAT, a novel analytical framework for assessing trail running performance based on Strava segmentation and GPX data. Such a validation process is essential when developing a new tool (Aithal & Aithal, 2020) to ensure that its content is appropriate for the intended purpose. Ten scientific and/or running experts were involved in the validation process for this study, in accordance with Gilbert & Prion (2016), who stated that content reviewers should be specialists in the field under study. The ideal number of contributors to an evaluation committee depends on the subject being analyzed. While three experts would suffice for a narrow domain; generally, 5 to 10 is preferable. On the other hand, involving more than 10 experts is probably superfluous, as Lynn (1986) indicates.

Our results show that the experts generally validated the TSAT for its relevance, both in terms of content and form. Our tool was deemed relevant and comprehensive. CVI analysis revealed that the majority of areas evaluated achieved a score of 0.8 or above.

The **TIC** emerged from this study as a robust framework for characterizing trail races, combining quantitative metrics and qualitative insights. Experts unanimously praised its comprehensive structure, which integrates key trail elements such as distance, elevation gain/loss (D+/D−), KmE, and technicity into a single, standardized profile. This synthesis is critical in trail running, where traditional performance models often fail to account for the interplay between physical exertion and environmental variability (Ehrström et al., 2018; Zimmermann et al., 2022). By translating complex course features into actionable data, the TIC bridges a gap between empirical analysis and practical utility, offering athletes, coaches and organizers a shared language for assessing race difficulty and planning race strategies. The main improvements made to the TIC concerned details on aid stations (number and segment), the technicity scale, segment categorization of segments and design.

The technicity scale was refined through experts' feedback by standardizing the terms, adding commitment and vigilance, and excluding weather conditions from level indicators. The experts received the scale very positively, noting it provided valuable additional information that complements the usual metrics such as total distance and elevation gain. By offering a structured representation of terrain difficulty, it contributes to a more comprehensive understanding of trail characteristics. However, the current version of the scale remains based on subjective assessment. Further studies are needed to establish its validity and reproducibility across different contexts and with more users. Moreover, the technicity of a trail or of specific segments can vary substantially depending on weather conditions, which are not currently accounted for in the TIC. Developing tools that can dynamically adjust to environmental variability and include weather conditions would be an exciting challenge for future research in trail running performance analysis.

Following expert recommendations, the TIC was refined by subdividing uphill and downhill segments according to slope intensity. In addition to the original 3% threshold, a second threshold at 10% was introduced to differentiate between moderate and steep gradients. This refinement is supported by several studies. Giandolini (2015) demonstrated that downhill slopes exceeding 10% substantially increase impact forces and neuromuscular fatigue in trail runners. Similarly, on steep uphill above 10%, walking becomes more energy-efficient than running, as shown by Minetti et al. (2002), Giovanelli et al. (2016), and McNeill Alexander (2002). This new categorization offers a more precise analysis of runners' adaptations across different terrain types.

In parallel, the TIC integrates a climb and descent classification model inspired by road cycling, based on a score calculated from the square of segment length and slope (Rodriguez-Marroyo et al., 2003). This method enables a graded assessment of elevation difficulty, categorizing segments from "uncategorized" to "out of category". Experts highlighted the relevance of this approach for events with substantial elevation gain, as it provides a standardized framework to quantify the intensity of vertical challenges. However, some noted that its applicability may be limited in flatter regions, such as Belgium.

The **TEA** component of the TSAT was also well received by the expert panel, who appreciated its ability to provide individualized insight into pacing strategies relative to course characteristics. The concept of ES, which normalizes performance across segments based on group averages and total race time, was considered a solid foundation for segment-level analysis. This method enables the calculation of %RSD, a central indicator for identifying effort distribution across various trail sections. Experts emphasized its relevance in understanding how runners adapt their pacing to different gradients or technical features. The ES and %RSD metrics introduced in the TSAT represent a significant advancement over existing pace analysis models such as Gradient-Adjusted Pace (GAP). While GAP normalizes speed based on theoretical elevation adjustments (Snyder & Farley, 2011), ES personalizes pace expectations by benchmarking each runner against their own performance level relative to a peer group. This individualization is critical in trail running, where effort distribution varies widely between athletes.

Trail runners exhibit distinct pacing strategies, with downhill velocity serving as a key performance discriminator. Elite runners maintain higher descent speeds through efficient eccentric control and technical skill (Giovanelli et al., 2016), while recreational athletes often brake excessively, increasing metabolic cost (Ehrström et al., 2018). These terrain-specific adaptations may reveal unique runner profiles such as "descenders" who excel on technical declines through neuromuscular efficiency, or "climbers" who preserve energy on ascents (Zimmermann et al., 2022). The classification of runners into "climbers," "descenders," or "balanced" profiles fits with this literature and was conceptually supported by experts, but two of them suggested adding a fourth category corresponding to those who excel on the flat. While it is very interesting to look at the relative speed of the runner in the flat segments, it cannot be included in an analysis based on %RSD differences between uphill and downhill segments. However, the TSAT offers an analysis by category of segment which is sufficient to identify a runner who is faster on flat sections or faster on hilly sections.

A major finding of our study is the uncertainty of experts regarding the thresholds used in the different classifications and interpretations of the TSAT. This largely stemmed from the provisional nature of these cut-off values, which were initially established through preliminary data analysis rather than population-derived benchmarks. The current thresholds serve as useful starting points for interpretation, but their arbitrary selection underscores an important limitation. To address this, future research should focus on establishing evidence-based standards through large-scale empirical studies of trail runner populations. Such work could employ percentiles or standard deviations to determine threshold limits (Harrell, 2015) or cluster analysis techniques (Jain, 2010) to identify natural groupings in pacing behavior across different terrain types. This data-driven approach would significantly enhance the TSAT's validity by replacing our current heuristic thresholds with statistically robust reference values that better reflect the diverse performance strategies observed in trail running.

The TSAT addresses a significant gap in trail running research, where traditional performance models fall short due to the sport's complex and variable nature (Chase & Hobbs, 2023; Ehrström et al., 2018). By incorporating terrain-specific segmentation and individualized pacing analysis, the tool aligns with recent calls for more context-sensitive performance metrics (Czegledi et al., 2023; Zimmermann et al., 2022). The use of the Strava segment analysis, whose chronometric measurements have been confirmed in some standardized situations (De Cock et al., 2023), indicates that the TSAT can be considered a reasonably reliable and accessible tool.

Strengths and limits

To our knowledge, the TSAT is the first tool that gives an analytical approach to performance and effort management for trail running. Although it does not provide a substitute for qualitative surveys carried out directly with runners, this tool can supplement them and provide a complementary quantitative view. In our case, this ability to make previously unmeasurable running activities measurable and comparable has led to the categorization of running practices. We would like to emphasize that this tool enabled us to analyze trail runs in greater detail while ensuring the reproducibility of this classification. In this way, the runs of each trail runner were distributed in a way that can be used for further analysis.

Our analysis of a runner's effort is based on the concept of ES, which can be challenged. The strength of this concept depends on the number of runners present in a trail race. The greater the number of participants, the greater the real dynamics of the race. In an event with low participation, potential interpretation biases may arise. That's why it's important to have a large database of participants.

Our tool could be improved by proposing an analysis of effort management based on the technical nature of the race. We have not yet proposed this option because, in the Belgian context, trail race environments have low technicity. In addition, researchers have observed that technicity is sometimes confused with slope. In fact, this concept is not always clear-cut, and we might ask: when does a slope become a technical element?

As mentioned by one of the experts interviewed, our analysis is based on the Strava application. We are dependent on this platform, with the risk of data loss in the event of changes or updates being made to it. Nevertheless, for the time being, Strava remains the most appropriate application for our segment analysis.

Although this study has partially validated the TSAT, the full validation process requires additional steps that remain important to ensure the tool's scientific robustness. Several CVI scores were below 0.8 for the thresholds chosen, indicating a need for further data-driven calibration.

Practical applications

Beyond contributing to theoretical advancements in sports performance research, the TSAT offers several concrete practical applications. Race organizers can use the tool to generate a detailed pre-race profile, allowing participants to anticipate the terrain, elevation difficulty, and technical features of the race. For athletes, the tool provides a means to study the course beforehand and later conduct a detailed post-race analysis of their pacing and effort management. Coaches and trainers can also benefit by using TSAT outputs to identify specific strengths and weaknesses in their athletes' performance (such as underperformance on steep uphill or inconsistent pacing) thereby informing more targeted and individualized training plans. The TSAT could serve as a resource for optimizing race strategies and monitoring performance.

Conclusion

This study aimed to validate TSAT using expert feedback through a mixed-methods approach. The results confirm the overall relevance and clarity of the tool, with strengths noted in its structure and conceptual design. Minor refinements were made, notably to the technicity scale and segment categorization, in response to the experts' suggestions. However, several proposed thresholds remain to be empirically validated, as the experts highlighted the need for data-driven confirmation.

Overall, this work represents an important first step in the development of a tool designed to enhance the understanding of trail effort management and performance. While the TSAT has shown strong initial validity, further studies are needed to refine thresholds and verify reproducibility and predictive value in diverse contexts. With further refinement, the TSAT could become an essential tool for optimizing training, race planning, and performance analysis in trail running.

Acknowledgments
The authors wish to thank all the runners, scientific and/or trail running experts for participating in this study.
Conflict of interest The authors state no conflict of interest.
Research funding None declared.

Research ethics

The study complies with the Declaration of Helsinki and has been approved by the ethics committee of the University of Liège (reference 2021/422).

Informed consent Procedures were explained to the participants who all gave their informed consent to participate.

Data availability

The data that support the findings of this study are available from the corresponding author Florence De Cock, upon reasonable request. Due to ethical and privacy considerations, the data are not publicly available as they contain sensitive information that could compromise the confidentiality of research participants.

References

Aithal, A., & Aithal, P. S. (2020). *Development and Validation of Survey Questionnaire & Experimental Data – A Systematical Review-based Statistical Approach* (SSRN Scholarly Paper No. 3724105). https://doi.org/10.2139/ssrn.3724105

Chase, A. W., & Hobbs, N. (2023). *Ultimate Guide to Trail Running : Everything You Need to Know about Equipment, Finding Trails, Nutrition, Hill Strategy, Racing, Avoiding Injury, Training, Weather, and Safety*. Rowman & Littlefield.

Czegledi, O., Beaucousin, S., Caballero-Julià, D., & Campillo, P. (2023). Digitalization in Trail Running : Digital Tools as Sustainable Outdoor Infrastructure. *Sustainability*, *15*(14), Article 14. https://doi.org/10.3390/su151411085

De Cock, F., Dardenne, N., Jockin, F., & Jidovtseff, B. (2023). Validity and reliability of STRAVA segments : Influence of running distance and velocity. *Journal of Human Sport and Exercise*, *18*(4). https://doi.org/10.14198/jhse.2023.184.05

Ehrström, S. (2020). *Analyse de la performance en trail courte distance : Déterminants physiologiques, spécificité de la sollicitation musculaire et stratégies d'optimisation*.

Ehrström, S., Tartaruga, M. P., Easthope, C. S., Brisswalter, J., Morin, J.-B., & Vercruyssen, F. (2018). Short Trail Running Race : Beyond the Classic Model for Endurance Running Performance. *Medicine & Science in Sports & Exercise*, *50*(3), 580-588. https://doi.org/10.1249/MSS.0000000000001467

Giandolini, M. (2015). *Gestion de l'impact et de la fatigue neuromusculaire en trail running* [Phdthesis, Université Jean Monnet - Saint-Etienne]. https://theses.hal.science/tel-01558928

Gilbert, G., & Prion, S. (2016). Making Sense of Methods and Measurement : Lawshe's Content Validity Index. *Clinical Simulation in Nursing*, *12*, 530-531. https://doi.org/10.1016/j.ecns.2016.08.002

Giovanelli, N., Ortiz, A. L. R., Henninger, K., & Kram, R. (2016). Energetics of vertical kilometer foot races; is steeper cheaper? *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *120*(3), 370-375. https://doi.org/10.1152/japplphysiol.00546.2015

Harrell, F. E. (2015). *Regression Modeling Strategies : With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-19425-7

Hoffman, M. D. (2010). Performance Trends in 161-km Ultramarathons. *International Journal of Sports Medicine*, *31*(01), 31-37. https://doi.org/10.1055/s-0029-1239561

ITRA Discover Trail Running. (s. d.). Consulté 19 août 2024, à l'adresse https://itra.run/About/DiscoverTrailRunning

Jain, A. K. (2010). Data clustering : 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

Koerner, H. (2019). *Le guide de l'ultrarunning : Préparez vos trail et ultra-trail*. Talent Sport.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382-385.

McNeill Alexander, R. (2002). Energetics and optimization of human walking and running : The 2000 Raymond Pearl memorial lecture. *American Journal of Human Biology: The Official Journal of the Human Biology Council*, *14*(5), 641-648. https://doi.org/10.1002/ajhb.10067

Minetti, A. E., Moia, C., Roi, G. S., Susta, D., & Ferretti, G. (2002). Energy cost of walking and running at extreme uphill and downhill slopes. *Journal of Applied Physiology*, *93*(3), 1039-1046. https://doi.org/10.1152/japplphysiol.01177.2001

Rodriguez-Marroyo, J., Garcia-Lopez, J., Avila, C., Jimenez, F., Córdova, A., & Villa, J. (2003). Intensity of Exercise according to Topography in Professional Cyclists. *Medicine and science in sports and exercise*, *35*, 1209-1215. https://doi.org/10.1249/01.MSS.0000074562.64053.4F

Santuz, A., Ekizos, A., Janshen, L., Mersmann, F., Bohm, S., Baltzopoulos, V., & Arampatzis, A. (2018). Modular Control of Human Movement During Running : An Open Access Data Set. *Frontiers in Physiology*, *9*. https://doi.org/10.3389/fphys.2018.01509

Scheer, V. (2019). Participation Trends of Ultra Endurance Events. *Sports Medicine and Arthroscopy Review*, *27*(1), 3. https://doi.org/10.1097/JSA.0000000000000198

Snyder, K. L., & Farley, C. T. (2011). Energetically optimal stride frequency in running : The effects of incline and decline. *Journal of Experimental Biology*, *214*(12), 2089-2095. https://doi.org/10.1242/jeb.053157

Stöhr, A., Nikolaidis, P. T., Villiger, E., Sousa, C. V., Scheer, V., Hill, L., & Knechtle, B. (2021). An Analysis of Participation and Performance of 2067 100-km Ultra-Marathons Worldwide. *International Journal of Environmental Research and Public Health*, *18*(2), Article 2. https://doi.org/10.3390/ijerph18020362

Yusoff, M. S. B. (2019). ABC of Content Validation and Content Validity Index Calculation. *Education in Medicine Journal*, *11*, 49-54. https://doi.org/10.21315/eimj2019.11.2.6

Zimmermann, P., Müller, N., Schöffl, V., Ehrlich, B., Moser, O., & Schöffl, I. (2022). The Energetic Costs of Uphill Locomotion in Trail Running : Physiological Consequences Due to Uphill Locomotion Pattern—A Feasibility Study. *Life*, *12*(12), Article 12. https://doi.org/10.3390/life12122070

Appendix I. Initial version of technicity scale

Score	Perceived level of technicity	Indicators	Examples
0	No technicity	No influence on the runner's running speed	A road jogging, flat running
1	Low level of technicity	Obstacles slightly reduce the runner's running speed	A forest path
2			
3			
4	Moderate level of technicity	Obstacles moderately reduce the runner's running speed	A rocky, fickle path
5			
6			
7	High level of technicity	Technicity greatly reduces running speed and requires the runner to adapt	A path filled with roots, an unstable and/or slippery ground surface (sand, gravel, mud, snow), ...
8			
9	Maximum level of technicity	The technical nature is such that it slows the runner down considerably, forcing him to walk or use his hands, requiring a great deal of adaptation on his part and rapid decision-making	Large obstacles requiring straddling, a very unstable and/or slippery ground surface (sand, gravel, mud, snow), ... → the hardest trail imaginable
10			

Appendix II. Final version of technicity scale

Score	Perceived level of technicity	Indicators	Examples
0	No technicity	No influence on the runner's running speed	A road jogging, flat running
1	Low level of technicity	Obstacles slightly reduce the runner's running speed	A forest path
2			
3			
4	Moderate level of technicity	Obstacles moderately reduce the runner's running speed	A rocky, fickle path
5			
6			
7	High level of technicity	Obstacles greatly reduce running speed, requiring adaptation and commitment on the part of the runner, as well as particular vigilance to avoid injury	A path filled with roots, an unstable ground surface (sand, gravel, mud, snow), ...
8			
9	Maximum level of technicity	The obstacles are such that they slow the runner down considerably, forcing him to walk or use his hands, requiring a great deal of adaptation and commitment on his part, as well as particular vigilance to avoid injury and rapid decision-making	Large obstacles requiring straddling, a very unstable ground surface (sand, gravel, mud, snow), off-path passages, ... → the hardest trail imaginable