Footwear Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tfws20

The test-retest reliability of ride perception during treadmill and outdoor running

Cristine E. Agresta, Jillian Peacock, Sam Guadagnino, Alicia Carmichael & Richard Gonzalez

To cite this article: Cristine E. Agresta, Jillian Peacock, Sam Guadagnino, Alicia Carmichael & Richard Gonzalez (2021): The test-retest reliability of ride perception during treadmill and outdoor running , Footwear Science, DOI: 10.1080/19424280.2021.1994021

To link to this article: https://doi.org/10.1080/19424280.2021.1994021

View supplementary material

Published online: 05 Nov 2021.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tfws20

FOOTWEAR SCIENCE
https://doi.org/10.1080/19424280.2021.1994021

The test-retest reliability of ride perception during treadmill and outdoor running

Cristine E. Agresta[a], Jillian Peacock[b], Sam Guadagnino[a], Alicia Carmichael[b] and Richard Gonzalez[b]

[a]Rehabilitation Medicine, University of Washington, Seattle, WA, USA; [b]Biosocial Methods Collaborative, University of Michigan, Ann Arbor, MI, USA

ABSTRACT

Subjective assessment of footwear experience during running is critical for recommendation and design. Recently, ride perception was quantified using an eight-dimension ratings scale. Ride varied by shoe and individual for treadmill running. However, it is unknown whether ride perception is a reliable measure in circumstances other than treadmill running and across sessions. The purpose of this study was to assess the intra-runner reliability of the previously developed ride rating 5-point semantic differential scale and the influence of surface on reliability ratings. We collected ride ratings in-person from nine runners in a controlled (lab) and semi-controlled (5 km outdoor paved path) environment and remotely from 177 runners in an uncontrolled collection. Post-run ride ratings were collected from each group following two separate running sessions on multiple surfaces using native shoes. We used intra-class correlation analysis to determine intra-runner reliability of ride ratings between separate runs in the same shoe on the same surface and in the same shoe on different surfaces. We used a reliability threshold of ≥0.70 as adequate. More than half of the in-person and slightly under half of the remote runners reached adequate reliability for the same surface runs. Reliability was lower for different surface runs for the remote runners but not the in-person runners. For in-person collection, firmness had adequate reliability during treadmill running, while awareness and sound level had good reliability outdoors. Weight and yield showed adequate and near adequate reliability for remote collections for same surface runs. Ride perception exhibits test-retest reliability in the assessment of subjective experience of footwear. Differences in ratings between surfaces indicate both the definition and construct of ride is captured with this subjective scale.

ARTICLE HISTORY
Received 17 July 2021
Accepted 12 October 2021

KEYWORDS
Footwear; subjective experience; remote; energy-return; comfort

Introduction

Running footwear recommendation models and product design have begun to incorporate measures of subjective experience. To date, the most common subjective measures assessed in relation to footwear are comfort and fit. Beyond increasing levels of consumer satisfaction, enhanced comfort is thought to reduce the risk of injury (Mundermann et al., 2002; Nigg, 2010; Nigg et al., 2015) and increase running economy (Luo et al., 2009), which together enhance training and performance capability. Thus, attempts have been made to improve methods for accurate assessment and quantification of comfort (Lindorfer et al., 2019). Despite improvements in the ability to assess comfort reliably, the utility of this measure for footwear design and selection for a specific shoe *feel* is still in question. Specifically, comfort may represent a first-order parameter where runners expect all footwear to be comfortable. Thus, assessing comfort may not be useful for assisting with footwear design or selection to elicit a specific running experience or facilitate a specific running purpose (e.g. speed work).

Ride perception, which is 'the feeling of the ground, shoe, and foot as the foot transitions through cycles of contact with the ground', (Agresta et al., 2020; Lam et al., 2018) has the potential to shape the experience of a run and may offer a new layer to subjective assessment of running footwear. That is, ride perception can offer a window into the experience that a runner seeks during a single run rather than a long-term result, like better performance or lower injury risk.

CONTACT Cristine E. Agresta cagresta@uw.edu
Supplemental data for this article is available online at https://doi.org/10.1080/19424280.2021.1994021.
© 2021 Informa UK Limited, trading as Taylor & Francis Group

Moreover, rating qualities of ride provides a more nuanced awareness compared to comfort. Initial studies (Lam et al., 2018; Schrödter et al., 2016) on ride assessment used the feeling of 'smoothness' to quantify subjective experience. However, smoothness likely represents another first-order parameter where almost all runners expect footwear to ride smoothly. More recent work (Agresta et al., 2020) developed a ride ratings scale to capture eight key qualities of ride perception that were identified by a combination of text mining and traditional qualitative analysis of interviews with lay running experts. This method involved querying runners about qualities related to energy return, ground feel, firmness, yield, awareness, sound level, speed, and weight. Findings suggested that ride is a multidimensional perception that varies across shoes and individuals.

Currently, ride perception has only been tested in a controlled lab setting (Agresta et al., 2020, Lam et al., 2018). While findings from previous studies suggest that assessing ride qualities immediately following a run provides accurate results compared to querying during the run (Agresta et al., 2020), there is no information on whether ride is reliable across running sessions. Moreover, ride perception is likely influenced by running surface, as the definition suggests. Thus, it is critical to understand how ride perception changes when running outside of a lab and not on a treadmill. The purpose of this study is to assess intra-runner and intra-ride quality psychometric reliability during treadmill and outdoor running. We hypothesized that runners would demonstrate moderate to high reliability for ride perception between run sessions on the same surface and that running on different surfaces would reduce intra-runner reliability, due to the influence of surface eliciting different ratings of ride qualities.

Methods

We examined data collected using a small in-person sample with controlled running trials in a lab setting and with semi-controlled running trials using a standardized outdoor paved path. To further investigate the ride ratings scale *in the wild* and increase the scale of real-world application, we examined data collected remotely from a large sample of participants. The purpose of both samples was to contextualize the reliability of the ride ratings between small-scale and real-world application. The fully remote data represents the least controlled setting.

Table 1. Ride qualities queried during and after each running trial.

	Ride quality	Anchor 1	Anchor 2
1	Speed	Slow	Fast
2	Yield	Rigid	Flexible
3	Ground feel	Smooth	Bumpy
4	Energy return	Responsive	Unresponsive
5	Awareness	Noticeable	Unobtrusive
6	Firmness	Soft	Firm
7	Weight	Light	Heavy
8	Sound level	Quiet	Loud

Qualities were rated on a 5-point scale with descriptors (anchor 1 and 2) at each end of the scale to anchor the perception spectrum.

Ride qualities

Eight themes, or ride qualities, were assessed during running trials: (1) energy return, (2) ground feel, (3) firmness, (4) yield, (5) awareness, (6) sound level, (7) speed, and (8) weight. Qualities were chosen based on previous work (Agresta et al., 2020). The eight resultant qualities were assessed on a 5-point semantic differential (bipolar) scale and anchored with specific descriptors (Table 1).

In-person collection

Subjects

Nine healthy distance runners (2 M) were recruited for this study through electronic flyers and word of mouth. Inclusion criteria were minimum running distance of 19 kilometres per week; age 18–55 years old; no lower extremity injuries in the previous 6 months; no current orthotics usage; and no previous experience in running footwear, sales, or coaching. Each participant provided written informed consent before involvement in the study. Data were collected following a protocol approved by the University Institutional Review Board on human subjects research. Participant demographics for in-person collection can be found in Table 2.

Test procedures

Prior to lab data collection, runners completed a digital survey to collect participant demographics and running characteristics. We defined ride as 'the

Table 2. In-person participant demographics and training characteristics.

Sample size	9 (7 Female)
Age (years)	29.6 (4.6)
Running characteristics	
Experience (years)	7.11 (5.01)
Typical training pace (minutes/km)	5.55 (0.64)
Training volume (km/week)	41.66 (43.13)
Type of runner	
Fitness/frequent runner	5
Jogger/recreational runner	3
Serious competitive runner	1
Not a runner	0

feeling of the ground, shoe, and foot as the foot transitions through cycles of contact with the ground'. This definition was chosen, in part, over previous definitions (Lam et al., 2018) to generalize to all runners regardless of foot strike pattern. A study investigator stated this definition to the participant prior to initial in-lab survey collection.

Participants ran on an indoor, instrumented treadmill (Bertec Corp., Columbus, OH, USA) for sixty seconds at 0% incline in their native running shoes. We selected sixty seconds of treadmill running as this is reflective of what most runners would experience when selecting shoes in-store. Treadmill speed was selected by the runner to represent a comfortable training pace. We instructed runners to select a speed that was representative of their typical running pace and one at which they could comfortably sustain for at least 45 minutes. Immediately following the treadmill running trial, runners completed a digital survey to quantify ride ratings of their native shoe for each ride quality. Runners were encouraged to run at the same comfortable speed during the outdoor run. A smartwatch (FenixPlus, Garmin, Inc., MN, USA) was used to record outdoor running speed. Immediately following the outdoor running trial, runners completed a digital survey to quantify ride ratings of their native shoe for each ride quality. Ride quality queries always appeared in the same order. If runners asked about how to interpret each ride quality, a study team investigator said that each descriptor can be interpreted as 'whatever that means to you'.

Runners returned for a second session where the same procedure was repeated. Treadmill speed for the second session was matched to the first session. In-person sessions were separated by an average of 17.4 ± 8.4 days.

Data analysis

The intra-runner reliability was determined by calculating the intraclass correlation coefficient (ICC) for each participant. The eight ride ratings provided one intra-runner reliability measure per participant, per condition. (i.e. comparing the ride ratings between the treadmill in different sessions and comparing the ride ratings between the outdoor paved path in different sessions).

The ICC indicates level of agreement between ratings at two timepoints. We used a threshold of 0.70 to indicate high or acceptable reliability that corresponds to previous work in footwear research (Hoerzer et al., 2016; Lindorfer et al., 2019). Data visualization of ICC values in relation to select runner and training characteristics was performed to identify potential trends in the data. Error bars represent plus/minus one standard error as defined from the F distribution; these error bars are scaled so that non-overlapping error bars are equivalent to statistical significance at alpha = 0.05 (uncorrected for multiple tests).

The influence of surface on ride was examined by calculating the ICC for each participant at each session (i.e. comparing the ride ratings between the treadmill and the outdoor paved path in the same session). To assess the intra-ride quality reliability, the ICC for each ride quality was examined for each condition (treadmill and outdoor paved path). Intra-ride quality reliability ratings were computed over runners so reflect inter-runner reliability assessments in contrast to the intra-runner reliability across the 8 ride ratings, which reflects test-retest consistency in how a single participant answered the 8 ride ratings across two sessions.

Remote collection

Subjects

One hundred seventy-seven (177) runners were recruited for this study through electronic flyers and word of mouth. Inclusion criteria were over the age of 18; live in the United States; and use a commercially available device or app that tracks, at a minimum, the distance, pace, and time of a run. Each participant provided digital, written informed consent before beginning the study. Data were collected following a protocol approved by University

Institutional Review Board on human subjects research.

Test procedures

Participation in the remote collections was entirely remote and accomplished through digital surveys. Runners were asked to complete up to four digital surveys. The initial survey (Runner Profile) collected the same participant demographics and running characteristics as the in-person collection. The definition of ride was stated in the survey and runners were asked to provide ride ratings of their shoes used during the run for each ride quality (Table 1).

Upon completion of the initial survey, runners were invited to participate three additional times and, to do so, were instructed to 'go for a run, and fill out a survey immediately after' with a minimum of 48 hours between surveys. Runners were also asked to rate comfort (least = 1 – most = 5) of their shoes. The additional three surveys collected the same run attributes and ride ratings of the shoes used for the run for each ride quality.

Data analysis

The analysed data was restricted to participants who reported at least two runs in the same shoe ($n = 123$, 35 M). These participants were further subset into two independent groups (no overlapping participants) where participants reported running on the same surface and in the same shoe for two runs (*same surfaces group*; $n = 83$) and participants who reported running on different surfaces for two runs (*different surfaces group,* $n = 40$). Each runner reported the surface(s) that the run was on by selecting categories from a list. During data analysis, these categories were reduced to five broad categories –pavement, trail, treadmill, track, and multi-surface. See Table 3 for the runner-selected surfaces that were included in each category.

Test-retest reliability between sessions and across ride qualities were determined using the same methods described above for the in-person collection. That is, the eight ride ratings were collapsed into one reliability measure per participant over their two reported runs to assess the intra-runner reliability. The intra-ride quality reliability was assessed by calculating ICC for each ride quality for each group (same surfaces group and different surfaces group). The error bars and the reliability threshold (0.70) were defined in the same way as for the in-person collection.

Runners were also queried about comfort of their shoe on a 5-point semantic differential (bipolar) scale. Comfort ratings were visualized by individual ride qualities to observe any trends in the data. We performed a one-way ANOVA to determine mean differences in ride ratings by comfort ratings. A Fisher's least significant difference test was used for pairwise comparisons by comparing overlapping error bars that are appropriately scaled.

Results

In-person collection

Fifty-six percent (5/9) of subjects reached or surpassed the 0.70 reliability threshold when running on treadmill surface and outdoor paved path (Figure 1). This value only dropped to 44% (4/9) for the session-to-session reliability threshold. Likewise, session-to-session reliability was slightly lower for outdoor paved path at 56% (5/9) (Figure 1). None of these differences are statistically significant.

Firmness (soft-firm) almost achieved acceptable reliability (ICC = 0.69) for treadmill running, while awareness (noticeable – unobtrusive) (ICC = 0.80) and sound level (quiet – loud) (ICC = 0.78) achieved acceptable reliability outdoors (Table 4). As expected, due to psychometric properties, single item reliabilities are lower than ride concept reliabilities. Subject demographics did not appear to be systematically associated with ride rating reliability (Figure 2).

A complete list of ICCs and respective Confidence Intervals (CI) are listed in Supplementary Material (Table S1).

Remote collections

Participant demographics for remote collection can be found in Table 5. Remote sessions in the same surfaces group were separated by an average of 9.0 ± 14.7 days and 10.1 ± 8.3 days in the different surfaces group. The type of running surfaces per run can be found in Table 6.

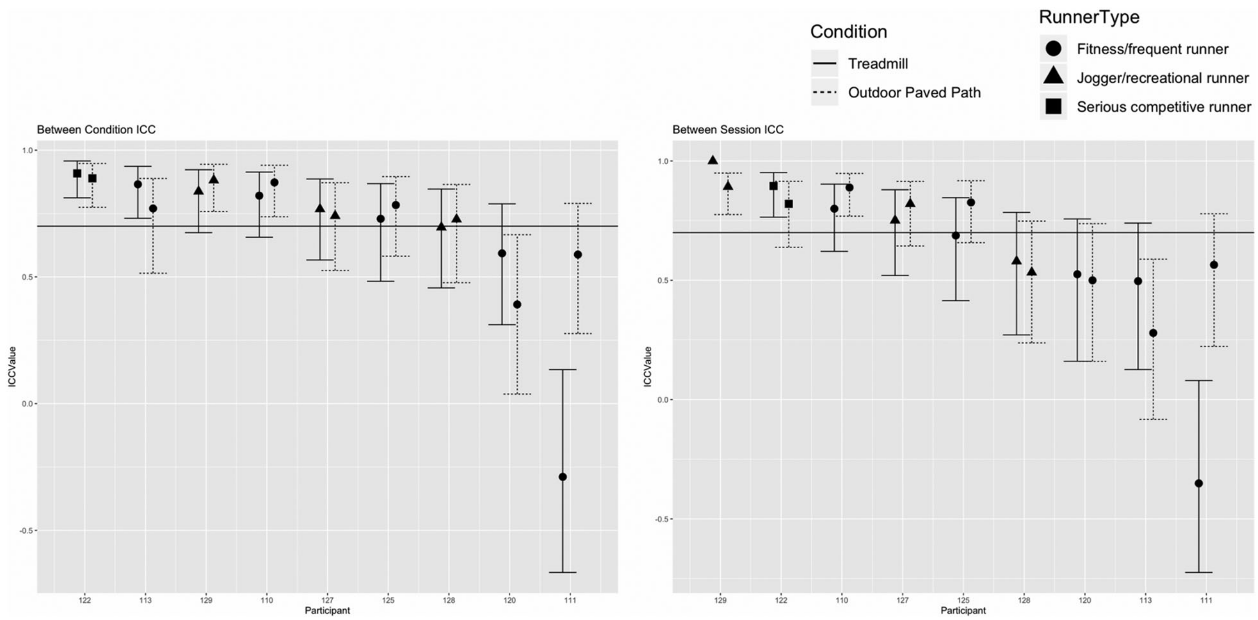

Figure 1. Intra-runner reliability for in-person collection between sessions and conditions. Reliability is represented by intra-class correlation coefficient (ICC) value on the y-axis. (*Left*) treadmill-to-treadmill reliability; (*right*) treadmill-to-outdoor paved path reliability. The solid lines represent values calculated from session one while dashed lines represent values calculated from session two.

Table 3. Surface categorization from remote collection responses.

Surface type	Participant selection
Pavement	Paved path, rural road, urban road
Trail	Dirt trail, grass, woodchips
Track	Indoor, outdoor
Treadmill	Home, gym
Multi-surface	(Paved path, rural road, urban road) AND (dirt trail, grass, woodchips)

In the remote collection, each participant was asked 'What kind of surface(s) was this run on? (Select all that apply.)' During data analysis, this selection was reduced to five categories as shown in the table below. If the participant chose surfaces that occurred in multiple categories (e.g. paved path and dirt trail), the surface type for the run was classified as multi-surface.

Table 4. Intra-class correlation coefficient (ICC) values for between-session reliability of each ride quality.

		Treadmill day to day ICC [LL, UL]	Paved path day to day ICC [LL, UL]
1	Speed (slow-fast)	−0.171 [−0.568, 0.234]	0.024 [−0.364, 0.388]
2	Yield (rigid-flexible)	0 [−0.407, 0.376]	0.349 [−0.028, 0.631]
3	Ground feel (smooth-bumpy)	−0.206 [−0.568, 0.182]	0 [−0.371, 0.360]
4	Energy return (responsive-unresponsive)	0.360 [−0.023, 0.640]	0.322 [−0.047, 0.610]
5	Awareness (noticeable-unobtrusive)	0.584 [0.277, 0.780]	0.800 [0.635, 0.898]
6	Firmness (soft-firm)	0.696 [0.465, 0.841]	0.273 [−0.048, 0.559]
7	Weight (light-heavy)	−0.405 [−0.752, 0.003]	0.512 [0.184, 0.735]
8	Sound level (quiet-loud)	0.042 [−0.368, 0.410]	0.789 [0.609, 0.893]

Columns denote different conditions for in-person collection (treadmill and outdoor paved path). *Note*: 68% confidence interval (1 standard error) is noted in brackets with the format, [LL, UL] LL: lower level; UL: upper level.

Forty-seven percent (39/83) of runners achieved or surpassed our selected threshold (ICC $\geq$0.70) for acceptable reliability the same surfaces group. This percentage was 40% (16/40) for the different surfaces group. These two percentages are not statistically different ($p = 0.46$ by Z test). Data visualisation did not reveal any observable trends for the influence of runner type, as self-categorised as serious competitive, recreational jogger, or frequent fitness runner by runners, on reliability for either surface (Figure 3). Runner demographics did not appear to influence ride rating reliability in runners (Figure 4).

Only weight (light-heavy) (ICC = 0.71) demonstrated acceptable reliability when running on the same surface with yield (rigid-flexible) almost

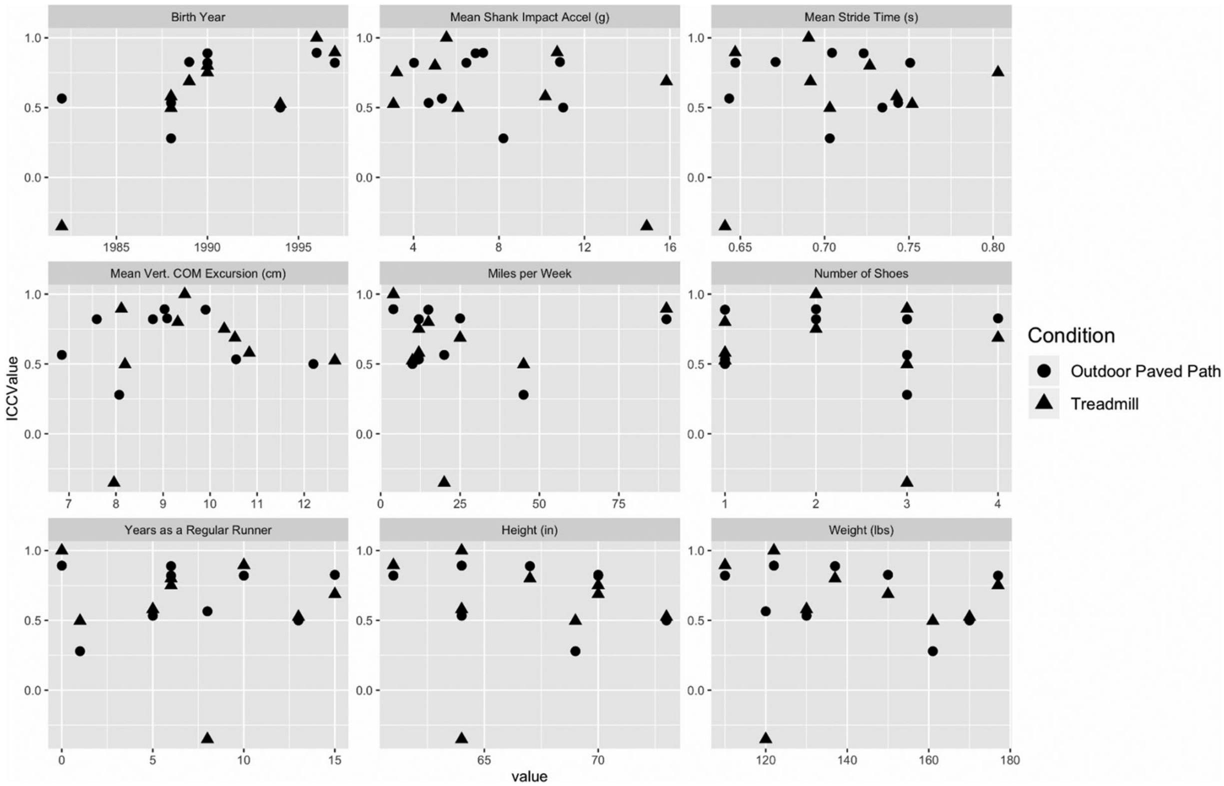

Figure 2. Runner demographic and training characteristic data in relation to intra-class correlation coefficient (ICC) value. No apparent trends were observed for select demographics or characteristics.

Table 5. Remote participant demographics and training characteristics.

	Same surfaces group	Different surfaces group
sample size	83 (61 Female)	40 (27 Female)
Age (years)	32.92 (12.89)	34.83 (12.83)
Running characteristics		
Experience (years)	7.27 (9.55)	8.68 (10.73)
Typical training pace (minutes/km)	4.9 (4.0)	5.7 (3.0)
Training load (km/week)	25.4 (28.4)	26.4 (24.0)
Type of runner		
Fitness/frequent runner	37	20
Jogger/recreational runner	37	14
Serious competitive runner	8	5
Not a runner	1	1

Table 6. Types of running surfaces for same surfaces and different surfaces groups.

	Same surfaces group	Different surfaces group
Multi-surface	8	24
Pavement	130	35
Track	4	2
Trail	10	11
Treadmill	14	8

Data is taken from remote collections only. *Note: Both runs in the same surfaces group occurred on the same surface. This table accounts for all runs.

achieving acceptable levels (ICC = 0.66). Running on different surfaces did not produce any acceptable reliability ride qualities.

Significant omnibus differences between ride scores and levels of comfort ratings were found for all ride qualities and comfort ($p \leq 0.05$) except firmness ($p = 0.06$). For these analyses, 'group' refers to the ride quality rating provided, and the analyses are the mean comfort in each such 'group'. Non-overlapping error bars within a ride characteristic are consistent with significant Fisher's significant difference tests (e.g. for speed, the means for comfort ratings 4 and 5 are significantly different from each other, both are different from the respective means for comfort ratings 2 and 3, but comfort ratings 2 and 3 are not different from each other).

A complete list of ICCs and respective CI levels for individual ride qualities by surface groups are listed in Supplementary Material (Table S2).

Discussion

The purpose of this study is to assess intra-runner and intra-ride reliability of ride ratings during treadmill and outdoor running. We hypothesized that runners would demonstrate moderate to high

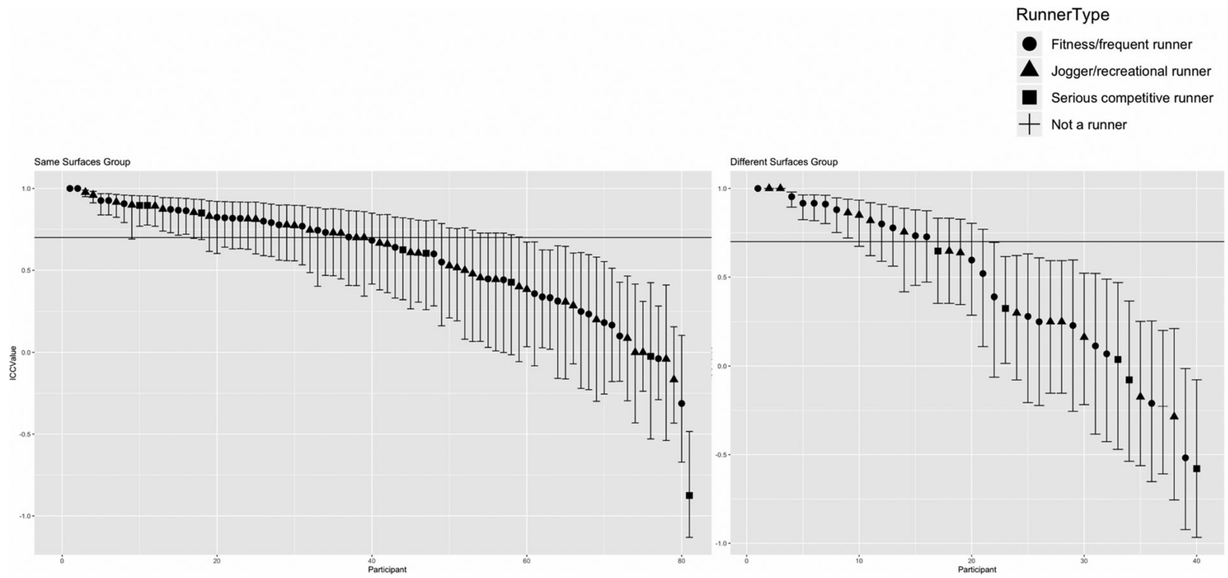

Figure 3. Intra-class correlation coefficient (ICC) values for runners in the remote collection sample. (*Left*) illustrates ICC values from the same surfaces groups; (*right*) illustrates ICC values from the different surfaces group. The acceptable reliability ICC threshold (0.70) is marked by the horizontal black line.

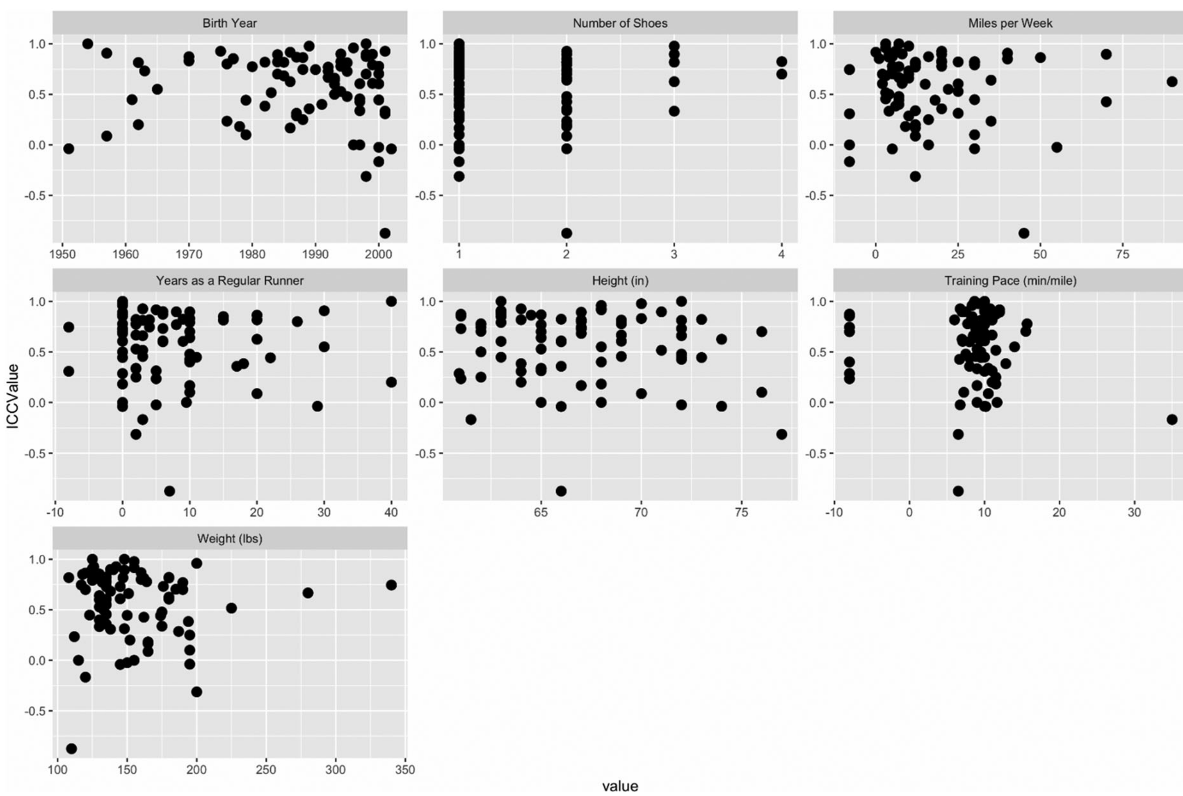

Figure 4. Runner demographic and training characteristic data in relation to intra-class correlation coefficient (ICC) value. No apparent trends were observed for select demographics or characteristics.

reliability for ride perception between run sessions on the same surface and that running on different surfaces would reduce intra-runner reliability and elicit different ratings of ride quality.

Intra-Runner reliability

In support of our hypothesis, we found that about 50% of runners in both in-person and remote collections reached the acceptable reliability threshold (ICC

≥0.70) with a slightly larger percentage (56%) of the in-person collections meeting the acceptability threshold for same surface running. This finding suggests that that overall ride perception appears to be as good, if not superior, to other subjective assessments of footwear. Previous work on reliability subjective measures found that between 30 to 60% of recreational runners had reliable comfort ratings as measured by the same ICC threshold value. Reliability was dependent on the assessment method, decreasing in reliability as complexity of the assessment method decreased (e.g. moving from VAS to yes/no query) (Hoerzer et al., 2016; Lindorfer et al., 2019).

It is imperative, as we move towards real-world collections like those illustrated by the remote collection in this study, that we understand what type of runners may give more reliable measures than others. Previous work has suggested selecting top raters (i.e. those with high reliability) to improve reliability of comfort ratings when assessing footwear. However, (Lindorfer et al., 2019) did not give an indication of what factors may point to top raters short of taking comfort measures in multiple sessions. Attempting to identify influential factors for reliable raters, we inspected years of running experience and training volume because they contribute to the amount of practice or total number of feedback cycles. In relation to ride perception, greater practice or feedback could help to refine the ability to perceive how a shoe feels while running. However, the relation between ride perception reliability and practice and/or feedback was not supported by the data. Visual inspection of experience and training volume did not reveal any trends to help explain why some runners had superior reliability. Likewise, there was no apparent trend between reliability and self-identified runner type (frequent fitness, serious competitive, or recreational jogger). It is possible that runners with higher reliability have better somatosensory perception. Future studies can employ tests (e.g. monofilament tests) to determine whether degree of perception influences reliability values). However, at this point, it remains unclear why some runners have higher ride perception reliability than others.

Intra-ride quality reliability

We found that specific ride qualities were more reliable depending on the running condition. Firmness was reliable on treadmill whereas sound level and awareness on paved path had acceptable reliability using the same threshold (ICC ≥0.70). However, this trend did not hold for the large remote collection. For the remote collection, we found that weight was reliable in the same surfaces group with yield almost reaching acceptable levels in the same group. No ride qualities were reliable in the different surfaces group. This contrast in findings may be because of our remote collection methodology. We included all runners who had at least two runs on the same surfaces, however, this surface did not have to only be treadmill or paved path. Only a small number of runs were on treadmill for runners in the remote collection. Interestingly, many participants ran on multiple surfaces within one run session, where multi-surface was categorised as selecting (Paved path, rural road, urban road) AND (Dirt Trail, Grass, Woodchips) on the post-run digital survey. Again, participant compliance limited the opportunity to collect GPS data to confirm the order and length of time spent on each surface.

The consistency of firmness and weight ride ratings during treadmill running may connect to previous findings of cushioning and mass perception, respectively. Isherwood et al. (Isherwood et al., 2021) found that runners were able to significantly distinguish between softer (Asker C ~47) and harder (Asker C ~57) shoes using a VAS scale with soft-hard anchors. While our ride qualities are not specifically linked to specific shoe features, runners may naturally intuit the qualities of firmness as the hardness feeling of a shoe. Likewise, Slade et al. (Slade et al., 2014) found that accuracy of mass perception from feet alone was about 30% for runners.

Findings from this study indicate that ride perception is unique from, but inter-related, to comfort. Some aspects of ride are more related to comfort than others. Importantly, ground feel (smooth – bumpy) has non-overlapping error bars for all levels of comfort (Figure 5). This indicates that the feeling of smoothness may substantially influence comfort rating. Likewise, weight (light – heavy) and sound level (quiet – loud) displayed non-overlapping error bars indicating these dimensions may also influence perception of comfort. Although, not as steep a distinction as ground feel, weight and sound level may

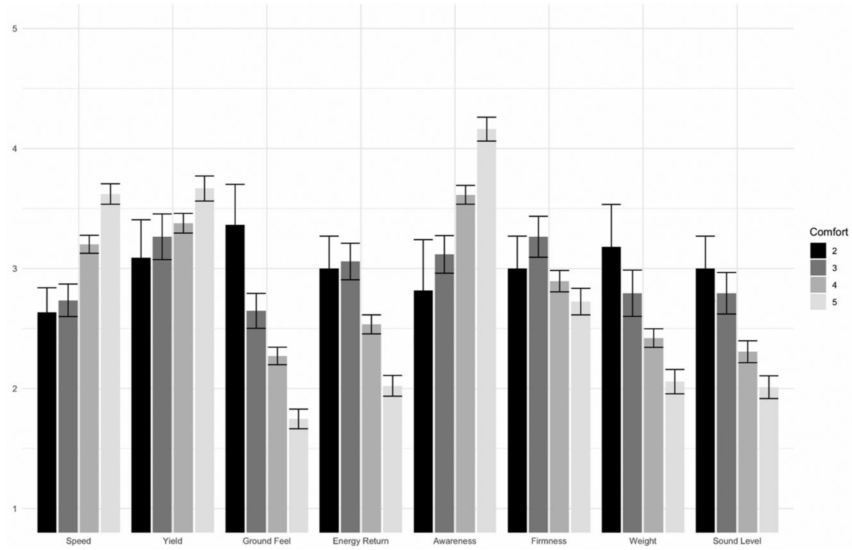

Figure 5. Mean comfort ratings and error bars for each comfort level in relation to ride ratings for remote collection sample. Comfort rating of 2 indicates most comfortable while 5 indicates least comfortable. Note that only ground feel, weight, and sound level had non-overlapping bars for the two levels indicating most comfortable (2 and 3).

offer dimensionality to comfort, particularly for outdoor running where reliability was high and differences between treadmill running were found. Differences in comfort perception, particularly for the most comfortable (2) and (3) rating levels, are not present for other ride qualities. This finding may help to explain why ride reliability is higher than comfort reliability. Typically comfort ratings are asked for shoe/foot regions. However, if the overall smoothness, weight, or sound level of the ride relate to comfort, it may be hard to accurately answer ratings based on shoe/foot regions. Runners may be unaware of the dimensional aspects they are intuiting to produce a subjective measure of comfort. Focussing their attention on specific ride qualities, in combination with overall comfort ratings, may help to increase the interpretation and reliability of both subjective measures. Additionally, yield and firmness display overlapping error bars across all comfort ratings. This finding suggests that these qualities are independent of comfort and represent ride as a unique perception. Moreover, the current findings support the notion that that yield and firmness heavily *weight* the dimensional perception of ride (Agresta et al., 2020).

Surface influence

Running surface influenced ride reliability. In partial support of our hypothesis, ride reliability was lower when running in the same shoe but on different surfaces. In this study, 40% of runners from the remote sample met the ICC threshold calculated for different surface runs compared to 47% on same surface runs. In contrast, the percent of runners from the in-person sample reaching acceptable reliability increased when comparing the different conditions (treadmill to outdoor paved path). The largest pairing of surfaces was multi-surface-to-pavement for the remote collection. One explanation for this discrepancy is that runners for the in-person sample ran on the same semi-controlled route and completed treadmill and outdoor paved path runs during the same data collection session while participants in the remote collection ran on the same surface but possibly not the exact route and not on the same day.

Post-hoc analysis of step, stride time, stride frequency, and vertical COM oscillation during both conditions from the in-person sample revealed no significant differences between conditions (treadmill or outdoor paved path) (See Supplementary Material, Table S3). Likewise, previous work has not found any consistent association between running biomechanics and subjective measures of comfort (Dinato et al., 2015; Keshvari et al., 2020; Lindorfer et al., 2020). Lam et al (Lam et al., 2017) found that smoother ride corresponded to a lower anterior-posterior velocity of the centre of pressure.

It is possible that daily state, or mindset, may influence a runner's perception of ride. Particularly, because differences between sessions were not well explained by differences in running biomechanics. However, more rigorous testing is needed to determine its merit. One area for potential exploration is to determine the influence, if any, of running cognitive strategies (Ogles et al., 1993) on ride perception and reliability.

Running surface influenced mean ride ratings between surfaces for the in-person sample, but not significantly. Post-hoc analysis (See Supplementary Material, Table S4) revealed that ride perception mean (group) values were relatively stable when running outdoors on a paved path compared to treadmill running. Unfortunately, we did not have enough treadmill-to-pavement pairings from the remote data collection to determine whether our findings from the in-person sample can be generalized to a larger population. The difference in intra-runner reliability between surfaces but not group values suggests that the overall ride quality may have a similar feel across runners, but the nuanced perception (degree and direction of change between surface) is individualized.

Limitations

Several limitations exist for this study. For in-person data collection, the treadmill was set to 0% incline because of hardware constraints rather than 1%, which is known to better replicate outdoor conditions (Jones & Doust, 1996). It is unknown how much, if any, the differences between surfaces were because of this. Secondly, while our inclusion criteria were selected so that runners would be able to comfortably run for the data collection periods without fatigue, some runners may have experienced fatigue, particularly during the outdoor run. We do not know the influence of fatigue on ride ratings. For the remote data collection, we condensed 'pavement' to include urban road, rural road, and paved path. We did not require participants to submit their GPS files with ride ratings. Because of this, compliance for GPS file submission was low and so we could not analyze differences between 'pavement' surfaces. Further, we don't have objective data from remote subjects to confirm that they ran immediately prior to completing the post-run ride ratings survey or specific attributes about their run (i.e. explicit terrain conditions). Future studies should aim to mirror external stimuli for both conditions and collect an export (or screenshot of a timestamp) from a wearable to confirm performance of the run.

Practical application

Ride perception appears to be as reliable a subjective measure as comfort. When assessing running footwear using ride perception, it is necessary to ask the runner which type of surface they intend to wear the shoe on and collect ride perception on that surface. Ride ratings gleaned from treadmill running may not be applicable to outdoor running on different surfaces, regardless of footwear type. If seeking comfort ratings for footwear, a 5-point semantic differential (bi-polar) scale querying ground feel (smooth-bumpy) and, perhaps, weight (light – heavy) or sound level (quiet – loud) may assist in interpretating results. As proposed in previous work (Agresta et al., 2020) and further supported here, queries of yield and firmness are influential dimensions that shape the unique perception of ride.

Future research should be focussed on refining measures of quality assurance for big (remote) data collection. For example, how to effectively capture GPS and/or smartwatch data along with ride ratings. The collection of such measures would allow for some statistical control of variables like surface/terrain, elevation, weather conditions, time of run (GPS-related measures) and biometric data that may help to interpret fatigue (e.g. change in heart rate and pace). It is important to note that use of consumer-grade devices can expand the scale and scope of remote data collection, but issues remain around how to best control for data quality and consistency across devices. Continued work is needed to establish the connections between specific shoe features and ride ratings as well as their interaction with the running surface. Finally, identification of training or runner characteristics that help to explain reliability would facilitate more precise interpretation of footwear assessments, especially when done remotely.

Conclusion

Assessing the subjective experience of ride appears to be as reliable, if not superior, to other subjective measures of footwear. The reliability of ride ratings is influenced by the runner and the running surface. Currently, the specific person-level factors that influence ride reliability are unclear. Overwhelmingly, the most popular running surface was pavement. The preponderance of pavement running surfaces along with finding that ride ratings and perception differs across surfaces highlights the need to assess ride with running surface(s).

Acknowledgments

The authors would like to thank Blayne Hettinga, PhD, for his thoughtful comments, feedback, and careful review of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

C. Agresta receives consulting fees from Diadora S.P.A. This study was supported by Adidas AG [AWD000151].

References

Agresta, C., Peacock, J., Carmichael, A., Nielsen, K. E., Zendler, J., & Gonzalez, R. (2020). The perception of ride is multidimensional for running footwear. *Footwear Science*, *12*(1), 15–24. https://doi.org/10.1080/19424280.2019.1706645

Dinato, R. C., Ribeiro, A. P., Butugan, M. K., Pereira, I. L. R., Onodera, A. N., & Sacco, I. C. N. (2015). Biomechanical variables and perception of comfort in running shoes with different cushioning technologies. *Journal of Science and Medicine in Sport*, *18*(1), 93–97.

Hoerzer, S., Trudeau, M. B., Edwards, W. B., & Nigg, B. M. (2016). Intra-rater reliability of footwear-related comfort assessments. *Footwear Science*, *8*(3), 155–163. https://doi.org/10.1080/19424280.2016.1195451

Isherwood, J., Rimmer, E., Fu, F., Xie, Z., & Sterzing, T. (2021). Biomechanical and perceptual cushioning sensitivity based on mechanical running shoe properties. *Footwear Science*, *13*(3), 221–211. https://doi.org/10.1080/19424280.2021.1913243

Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences*, *14*(4), 321–327. https://doi.org/10.1080/02640419608727717

Keshvari, B., Schubert, N., Senner, V., & Mitternacht, J. (2020). Perception of running shoe cushioning: Objective and subjective measurements in short-distance running. *Proceedings*, *49*(1), 121. https://doi.org/10.3390/proceedings2020049121

Lam, C. K.-Y., Mohr, M., Enders, H., Nigg, S., & Nigg, B. (2017). Subjective and biomechanical assessment of 'ride' during running. *Footwear Science*, *9*(sup1), S42–S43. https://doi.org/10.1080/19424280.2017.1314327

Lam, C. K.-Y., Mohr, M., Nigg, S., & Nigg, B. (2018). Definition and quantification of 'ride' during running. *Footwear Science*, *10*(2), 77–76. https://doi.org/10.1080/19424280.2018.1453872

Lindorfer, J., Kroll, J., & Schwameder, H. (2019). Comfort assessment of running footwear: Does assessment type affect inter-session reliability? *European Journal of Sport Science*, *19*(2), 177–185. https://doi.org/10.1080/17461391.2018.1502358

Lindorfer, J., Kroll, J., & Schwameder, H. (2020). Does enhanced footwear comfort affect oxygen consumption and running biomechanics? *European Journal of Sport Science*, *20*(4), 468–476. https://doi.org/10.1080/17461391.2019.1640288

Luo, G., Stergiou, P., Worobets, J., Nigg, B., & Stefanyshyn, D. (2009). Improved footwear comfort reduces oxygen consumption during running. *Footwear Science*, *1*(1), 25–29. https://doi.org/10.1080/19424280902993001

Mundermann, A., Nigg, B., Stefanyshyn, D., & Humble, R. N. (2002). Development of a reliable method to assess footwear comfort during running. *Gait & Posture*, *16*(1), 38–45.

Nigg, B. M. (2010). *Biomechanics of sport shoes*. University of Calgary.

Nigg, B., Baltich, J., Hoerzer, S., & Enders, H. (2015). Running shoes and running injuries: mythbusting and a proposal for two new paradigms: 'preferred movement path' and 'comfort filter. *British Journal of Sports Medicine*, *49*(20), 1290–1294. https://doi.org/10.1136/bjsports-2015-095054

Ogles, B. M., Lynn, S. J., Masters, K. S., Hoefel, T. D., & Marsden, K. A. (1993). Runners' cognitive strategies and motivations: Absorption, fantasy style, and dissociative experiences. *Imagination, Cognition and Personality*, *13*(2), 163–174. https://doi.org/10.2190/p1dg-kkkj-agd0-a216

Schrödter, E., Brüggemann, G.-P., Hamill, J., Rohr, E., & Willwacher, S. (2016). Footwear-related variability in running. *Footwear Science*, *8*(1), 23–31. https://doi.org/10.1080/19424280.2016.1142002

Slade, S. J., Greenya, J. G., Kliethermes, C. L., & Senchina, D. S. (2014). Somatosensory perception of running shoe mass. *Ergonomics*, *57*(6), 912–920. https://doi.org/10.1080/00140139.2014.904009