Journal of Urban Mobility 4 (2023) 100066

Contents lists available at ScienceDirect

Journal of Urban Mobility

journal homepage: www.elsevier.com/locate/urbmob

I bet you feel safe! assessing cyclists' subjective safety by objective scores

Stefan Fuest [a,*], Mariana Batista [b], Frauke Luise Berghoefer [c], Morten Flesser [b], Bhagya Shrithi Grandhi [b], Felix Spühler [d], Monika Sester [a], Mark Vollrath [c]

[a] Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany
[b] Institute of Transportation and Urban Engineering, Technische Universität Braunschweig, Germany
[c] Department of Traffic and Engineering Psychology, Technische Universität Braunschweig, Germany
[d] Decision Support Group, Technische Universität Braunschweig, Germany

ARTICLE INFO

Keywords:
Cycling safety
Risk perception
Objective safety
Bike score
Cycling infrastructure

ABSTRACT

Feeling safe is a major issue for cyclists, and some potential cyclists are still deterred from using the bicycle because they feel too unsafe. Assessing the subjective safety of existing cycling infrastructures and locations can be done by questionnaires that show pictures of infrastructures and ask participants for their safety ratings. However, future cycling infrastructures should also be evaluated as safe even before they are implemented. Therefore, it is desirable to have a method that is able to predict safety from infrastructural information. This study aims to propose two different ways for such a method and to test both ways in a use case. We first developed two scores, namely the *Repertory Grid (RG) Score* and the *FixMyBerlin (FMB) Score*, which predict subjective safety from objective environmental information but use different data bases and different methodologies. In a second step, we validated these scores by comparing them to questionnaire ratings that evaluated cyclists' subjective safety at 20 locations in the city of Braunschweig, Germany. Finally, we compared the two scores as well as the questionnaire ratings with objective safety measures, namely crash statistics, at the respective locations. The results show that the RG Score has a moderate agreement and the FMB Score has a fair agreement with the questionnaire ratings. All methods agree on the overall safety evaluation of various cycling facilities. However, the RG Score showed less variance in the safety ratings, whereas the FMB Score rated most locations more unsafe than the participants in the questionnaire. Interestingly, neither the scores nor the questionnaire ratings could sufficiently deduce the occurrence of a crash at one of the locations. The findings strengthen the importance of subjective safety as a construct independent of objective safety. Furthermore, they provide insights into aspects of subjective safety that can easily be measured by objective scores, and into aspects that are important for cyclists but were not yet covered by the scores. This study, therefore, provides a basis for future considerations and future evaluation methods to assess the subjective safety of cyclists.

Introduction

Designing and implementing safe infrastructure is essential to promote cycling as an attractive mode of transport (Hull & O'Holleran, 2014; Pucher & Buehler, 2008; Dill, 2009). Since feeling safe is a crucial factor to encourage and keep people cycling, cycling infrastructure should not only ensure low levels of crashes but should also be perceived as safe (Hull & O'Holleran, 2014; Pucher & Buehler, 2008). Yet, subjective safety is often neglected when planning or designing cycling facilities (McLeod et al., 2020), although previous studies indicate that subjective safety is highly influenced by the design of the infrastructure (Berghoefer & Vollrath, 2022; Gössling & McRae, 2022; von Stülpnagel & Binnig, 2022). Instead of implementing infrastructure and then assessing subjective safety via questionnaires, it would be very efficient to be able to predict subjective safety based on the objective features of the infrastructure.

This paper aims to present objective approaches to reliably predict cyclists' subjective safety in any cycling facility. Therefore, we developed two different scoring methods to assess cyclists' subjective safety, the Repertory Grid (RG) Score based on an interview study (Berghoefer & Vollrath, 2022) and the FixMyBerlin (FMB) Score based on the FixMyBerlin survey (FixMyBerlin, 2020a). These scoring methods were applied to compute cyclists' predicted subjective safety ratings in four case study areas in Braunschweig, Germany. To validate these results,

* Corresponding author at: Appelstraße 9a, 30167 Hannover, Germany.
E-mail address: stefan.fuest@ikg.uni-hannover.de (S. Fuest).

https://doi.org/10.1016/j.urbmob.2023.100066
Received 21 April 2023; Received in revised form 23 June 2023; Accepted 21 September 2023
Available online 28 September 2023

we used an online questionnaire to collect safety ratings of 20 locations from the case study areas. In addition, the RG Score, the FMB Score, and questionnaire results were compared to reported crashes in these locations.

By introducing both scoring methods, this paper provides valuable insights into aspects of subjective safety that can be measured according to the design characteristics of different cycling facilities. Furthermore, the validation of the scoring methods with the questionnaire ratings presents information on the goodness-of-fit of the scores and the suitability of different infrastructural attributes for predicting subjective safety.

Literature review

The decision to ride a bicycle as a means of transport is subject to various factors. Besides factors such as the weather, the physical condition of the cyclist, the attractiveness and comfort of the cycling infrastructure, or the destination with its choice of route, it is primarily a matter of objective and subjective safety. To feel safe is an essential factor when choosing the bicycle as a mode of transport (Aziz et al., 2018). In literature, various tools, models, indices, and scores have been developed that assess cyclists' safety (Castañon & Ribeiro, 2021). Both objective and subjective approaches can be found in the literature to examine three concepts of safety, namely *actual, perceived*, and *inferred* safety. Actual or objective safety is based on direct objective measures, such as crash data, while perceived or subjective safety is examined by user surveys. Inferred safety explores safety via indirect measures that indicate unsafe situations, such as the distance between cyclists and overtaking motor vehicles (Daraei et al., 2021).

To assess actual or objective safety, studies use objective measures including data on infrastructure, network, traffic, land use, and crashes (e.g., Allen-Munley & Daniel, 2006; Aziz et al., 2018; Daraei et al., 2021). These quantitative data help to examine crash hotspots and shortcomings of the infrastructure and to develop realistic traffic models. However, cyclists' safety may be biased when assessed by objective measures only. First, because crash statistics often underestimate the actual risk, as they rarely include near-misses or crashes that were not reported to the police. Second, because cyclists usually are not aware of crash statistics, it is the cyclists' perception of safety rather than the objective safety that makes an infrastructure more or less attractive for cycling (Gössling & McRae, 2022).

Therefore, more and more studies focus on subjective safety assessed by questionnaires or interviews (Hardinghaus & Papantoniou, 2020; Desjardins et al., 2021; Berghoefer & Vollrath, 2022). In contrast to objective data, questionnaires can provide safety evaluations of existing infrastructure as well as of hypothetical or future cycling infrastructures. For instance, a comprehensive study, which developed a bike barometer to investigate adolescents' subjective perception of safety on routes, shows the potential of assessing existing infrastructure with citizen science (Storme et al., 2022). Furthermore, questionnaires can be conducted in an experimental way allowing conclusions about cause-effect relationships in the safety evaluation.

Although the methods and databases differ between studies on objective and subjective safety, they often agree on their results, that is, on factors that increase or decrease cyclists' safety. Dedicated cycle tracks that separate cyclists from motor traffic increases both the objective and subjective safety (von Stülpnagel et al., 2022) and the more the cycling facility is separated the more it is perceived to be safe and preferred by cyclists (Berghoefer & Vollrath, 2022; Hardinghaus & Papantoniou, 2020). Similarly, streets with less traffic volume and lower speed limit for the motor traffic as well as residential streets are found to be safer objectively as well as subjectively (Desjardins et al., 2021; von Stülpnagel et al., 2022; Winters et al., 2012).

However, in some aspects objective and subjective safety seems to differ. Cyclists tend to overestimate the safety of unpaved multi-use paths, of intersections between a main street and a side street, and when cycling against the travel direction of motor vehicles (von Stülpnagel et al., 2022; Winters et al., 2012), whereas they underestimate the safety of separated cycle tracks (Winters et al., 2012).

Although the evaluation of objective and subjective safety has been addressed by several researchers, we identified a few gaps in literature that have not yet been addressed extensively. As discussed in the review of Castañon and Ribeiro (2021), there is a need for validation processes. Apart from that, most existing studies are limited to one single approach and a combination of both objective and subjective indicators is required to properly assess safety. Therefore, in this paper, we propose two methods to predict subjective safety based on objective measures. The resulting two scores were validated using an online questionnaire about the subjective safety of different local areas in Braunschweig, Germany.

Scoring methods

Two scoring methods were developed to estimate cyclists' safety based on subjective safety perceptions in an objective manner: the Repertory Grid (RG) Score and the FixMyBerlin (FMB) Score. It is important to highlight that we consider four main types of cycling infrastructure in the scoring methods, namely cycle tracks, cycle lanes, advisory lanes, and streets with mixed traffic. A cycle track is a dedicated cycling facility that runs adjacent to, but separated from motor and pedestrian traffic, usually on the level of the sidewalk. Cycle lanes run on the street level and are separated by continuous lane markings that motor vehicles are not allowed to pass. The advisory lane is separated from the motor traffic by dashed lines and motor vehicles are allowed to use this space if not endangering cyclists. Streets with mixed traffic are usually streets with low speed limits, where cyclists and motor vehicles are expected to share space. For this kind of facility, we consider different street types, namely side street, traffic-calming zone, shared space, and bicycle boulevard.

Scoring method I: repertory grid (RG) score

The Repertory Grid (RG) Score is based on the results of a previously conducted interview study that uses photographs of different infrastructural characteristics such as various cycling facilities or traffic volumes to investigate how cyclists feel when exposed to different cycling infrastructures (Berghoefer & Vollrath, 2022). The used infrastructure characteristics comprise 15 elements, for which safety ratings have been provided by participants, namely: joint sidewalk and cycle track, main street, side street, separate cycle track, adjacent cycle track, cycle lane, advisory lane, bicycle boulevard, shared space, park, traffic signal, cobblestone, asphalt, high traffic volume, and low traffic volume. An overview of the photographs used for the different characteristics is provided in Appendix A. The interview study resulted in individual safety statements of how participants describe and evaluate the depicted infrastructure characteristics. Considering these statements, participants then rated the infrastructure characteristics on a scale from 1 (*very positive* pole) to 5 (*very negative* pole). To ensure comparability with the ratings resulting from the FMB Score and the questionnaire, the original ratings were reversed to a scale ranging from 1 (*very unsafe*) to 5 (*very safe*). For defining our scoring method, the safety statements collected in Berghoefer and Vollrath (2022) were assigned to different safety categories to account for differences of statements regarding addressed safety aspects. The safety categories are: A) general safety, B) cycling infrastructure, and C) other traffic participants.

Fig. 1 visualizes the median rating values of each addressed infrastructure characteristics for the three defined safety categories. Higher values indicate safe evaluations while lower values indicate unsafe evaluations. The characteristics on the x-axis are shown in descending order for the safety category *general safety*. We can observe that the safety estimation clearly differs among the different infrastructure characteristics. Regarding street types, it becomes apparent that cycling infrastructure which is clearly separated from motor traffic, such as

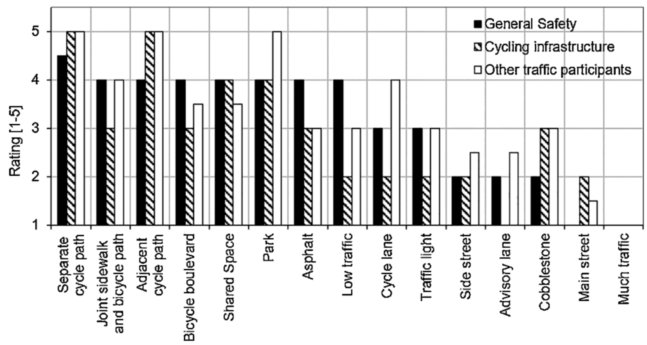


Fig. 1. Median ratings for infrastructure characteristics and the three safety categories. Higher ratings relate to a safer evaluation, lower ratings to more unsafe evaluations. The characteristics are ordered from safe to unsafe general safety.

separate and adjacent cycle tracks, are generally rated safer. However, cycling infrastructure that is directly adjacent to lanes dedicated to motorized traffic, such as advisory lanes or cycle lanes, are rated as less safe. Similarly, cycling on infrastructure that is shared with faster motorized traffic such as on main streets or side streets is perceived as less safe. Concerning these types of streets, a high motorized traffic volume (much traffic) is rated as less safe as compared to a low traffic volume (low traffic). For some of the infrastructure characteristics, we can also observe larger differences between safety categories. The cycle lane, for example, achieves a high rating in the category *other traffic participants* which indicates a perception of higher safety with respect to the risks posed by other traffic participants, but regarding the type of the *cycling infrastructure*, the cycle lane is rated unsafe.

For the RG Score, each street segment had a maximum of four rating values related to different groups of infrastructure characteristics: street types, traffic signal, surface, or traffic volume. To identify the final scores, we calculated the median value among all ratings assigned to a street segment. Detailed insights into the structure of the attribute table can be found in Appendix B.

Scoring method II: FixMyBerlin (FMB) score

The FixMyBerlin (FMB) Score is based on the FixMyBerlin survey, part of a project from FixMyCity to support transforming Berlin into a cycling city. In 2019, an online survey called "Street check-up", carried out with the Berlin newspaper *Tagesspiegel*, collected ratings from about 21.000 participants on the subjective safety of 1.900 photorealistic illustrations from the cyclists' perspective. These illustrations varied in characteristics of the cycling infrastructure on street segments, such as path width, surface, physical barriers, and on-street parking (FixMyBerlin, 2020a).

The survey included three main types of cycling infrastructure: cycle track, cycle lane, and streets with mixed traffic, as shown in Fig. 2. For each type, there were numerous variations, including but not limited to the type of cycle infrastructure, the presence of different barriers between the cycle track and sidewalk or street, colored surface, different levels of traffic and speed limit, markings designating bicycle boulevard or traffic-calming zone, and varied width. Since the FixMyBerlin survey

Fig. 2. Example of the photorealistic illustrations for a cycle track, mandatory cycle lane, and mixed traffic scenarios (FixMyBerlin, 2020b).

also aims to understand which design measures could increase subjective safety, some of the survey's pictures also presented elements unfamiliar to the German street configuration, such as bicycle boulevards with colored surfaces. Each participant was randomly assigned to ten photorealistic illustrations to rate the particular scene as *unsafe, rather unsafe, rather safe* or *safe*, resulting in an ordinal scale ranging from 0 (*unsafe*) to 3 (*safe*).

To associate the cyclist's subjective safety with different configurations of cycle tracks, cycle lanes, and mixed traffic conditions currently in place, we developed a simplified scoring method based on the results of this survey (FixMyBerlin, 2020a). Therefore, we reduced the original data set to focus on photorealistic illustrations with ratings for relevant design elements available on OpenStreetMap (OSM) and Google Earth. Consequently, not all but only some generic characteristics from the original survey were considered for the different types of cycling facilities (see Parameters in Table 1). In addition, we aggregated the survey results to consider all ratings, regardless of the participants' demographics or cycling experience. For cycle lanes, we also did not distinguish between advisory and mandatory cycle lanes. Instead, we considered main streets with markings, no markings, or colored surface.

In the FixMyBerlin survey, cyclists' subjective safety was evaluated using a 4-point Likert scale, considering various characteristics of cycling infrastructure. To compute the FMB Score for cycle lanes, cycle tracks, and infrastructure with mixed traffic, ordinal logistic regressions were then employed (Agresti, 2002). These regressions estimated models that quantified the influence of design elements on the likelihood of receiving each rating from 0 to 3. The coefficients in Table 1 provide insights into the magnitude of this influence with respective p-values, calculated by comparing the t value against the standard normal distribution. The goodness of fit of each ordinal logistic regression was evaluated through a likelihood-ratio test (Agresti, 2002). In all cases, the null hypothesis that the model without predictors is as effective as the one with them was rejected (p-value < 0.001), indicating that the predictors significantly contribute to the estimation process.

To compute a score for a particular cycling facility, one can consider the coefficients in Table 1 to estimate the probabilities associated with each rating taking into account the characteristics of the cycling infrastructure. The final score is defined based on the rating with the highest probability, representing the rating that had the highest likelihood of being chosen by the participant for that particular street configuration. It is important to ponder that the results can only reflect the influence of these design elements to a rather limited extent. Apart from considering only some of the attributes presented in the survey, our approach did not consider many other factors influencing the participants' perceptions, such as demographic and cycling background.

Case study

Cycling infrastructure in Braunschweig

Braunschweig is a city in the northern part of Germany with around 250.000 residents (City of Braunschweig, 2022a). Focusing on bicycle friendliness, Braunschweig reached rank 5 in the "Bike-Friendly Cities Rating 2020″ compared to cities with a similar size in Germany (German Cyclists Association Braunschweig, 2020). This bicycle-friendliness is reflected in the city's modal split, as around 21 % of the city's population use bicycles as means of transport (Stadt Braunschweig, 2022). Additionally, the city council and various bicycle and mobility associations aim to extend the existing cycling infrastructure by a special velo route network (German Cyclists Association Braunschweig, 2020).

For the case study, we aimed for areas that vary in land use, in their cycling facilities, their crash frequency as well as in their general bikeability. To assess the latter, we adapted a Bike Score that rates each road section relative to all road sections in a given area (Bike Score, 2022). This score is based on a Bike Lane Score which rates different

Table 1
Derived coefficients of FMB-scoring method.

	Infrastructure with mixed traffic								
	Side street			Bicycle boulevard	Traffic-calming zone				
Parameters	N	Coeff.	P	N	Coeff.	p	N	Coeff.	p
Intercept	8512			8498			8356		
0|1		−0.116	.127		−1.314	.001		−0.616	< 0.001
1|2		1.548	< 0.001		0.516	< 0.001		1.162	< 0.001
2|3		3.290	< 0.001		2.476	< 0.001		3.139	< 0.001
Width									
Narrow (< 6 m)	2880	0.000		2836	0.000		2763	0.000	
Wide (≥ 6 m)	5632	0.639	< 0.001	5662	0.631	< 0.001	5593	0.508	< 0.001
Traffic direction									
Two-way	2769	0.000		2904	0.000		2760	0.000	
One-way	2822	0.524	< 0.001	2797	0.381	< 0.001	2756	0.257	< 0.001
Contraflow-one-way	2921	−0.021	.744	2797	−0.271	< 0.001	2840	−0.262	< 0.001
Parking									
None	2873	0.000		2827	0.000		2779	0.000	
One-sided	2823	−0.432	< 0.001	2869	−0.418	< 0.001	2752	−0.246	< 0.001
Two-sided	2816	−0.398	< 0.001	2802	−0.446	< 0.001	2825	−0.296	< 0.001
Traffic flow									
Normal	7112	0.000	< 0.001	7004	0.000	< 0.001	6990	0.000	< 0.001
Car-free	1400	1.292	< 0.001	1494	1.768	< 0.001	1366	1.417	< 0.001
Null log-likelihood		−10,633			−11,214			−10,914	
Model log-likelihood		−10,153	< 0.001		−10,374	< 0.001		−10,377	< 0.001

	Separated infrastructure						
	Cycle lane				Cycle track		
Parameters	N	Coeff.	p	Parameters	N	Coeff.	p
Intercept	37,865			Intercept	135,800		
0|1		−0.822	< 0.001	0|1		−1.715	<0.001
1|2		0.910	< 0.001	1|2		0.142	<0.001
2|3		2.805	< 0.001	2|3		1.920	<0.001
Width (m)	37,865	0.775	< 0.001	Width (m)	135,800	0.883	<0.001
Traffic flow				Environment left side			
Normal	18,983	0.000		Street	56,558	0.000	
High	18,882	−0.090	<0.001	Greenery	31,714	0.122	<0.001
Speed limit				Parking	47,528	−0.055	<0.001
50 km/h	18,897	0.000		Environment right side			
30 km/h	18,968	0.184	<0.001	Sidewalk	119,800	0.000	
Markings				Greenery	16,000	0.511	<0.001
Line markings	17,306	0.000		Null log-likelihood		−48,990	
No markings	17,323	−0.696	<0.001	Model log-likelihood		−43,718	<0.001
Color marking	3236	0.474	<0.001				
Physical Protection							
None	34,283	0.000					
Any	3582	1.662	<0.001				
Null log-likelihood		−152,565					
Model log-likelihood		−141,970	<0.001				

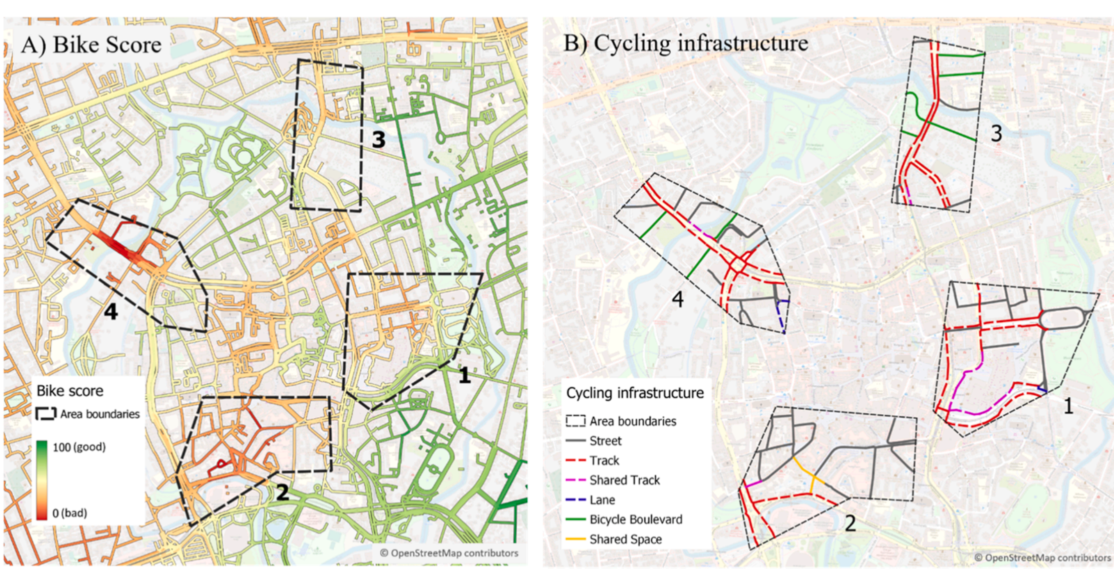


Fig. 3. Bike Score (A) and cycling infrastructures (B) in the four case study areas in Braunschweig.

infrastructure elements (OpenStreetMap, 2022), a Hill Score (Federal Agency for Cartography & Geodesy, 2022), a Destinations and Connectivity Score that rates the accessibility of amenities like restaurants or public buildings (OpenStreetMap, 2022), and a Crash Score with the number of crashes (Statistische Ämter des Bundes und der Länder, 2022).

Due to partially missing data sources, parts of the data generation process had to be done manually. Hence, in this work, we have refrained from applying our methodology to the entire urban area. Instead, we focus on four areas in the inner city of Braunschweig that met the above mentioned requirements and provide a large variety of different cycling facilities concentrated within a small space.

Fig. 3A shows the Bike Score for the inner city of Braunschweig including the four case areas. It can be observed that areas 1 and 3 depict a slightly better Bike Score rating which is mainly due to the ratings provided by the Bike Lane Score and the Crash Score. In addition, it should be noted that the relatively good scores are mainly found in park-like areas, as all sub scores are rated relatively good for the street sections. Due to the relative rating of the Bike Score, the other street sections are automatically rated worse.

Fig. 3B depicts the cycling facilities in the four case study areas. All main streets in the four areas provide a dedicated cycle track. The side streets, which are either designed as bicycle boulevard or without providing any cycling facility, require cyclists to share the street with motor traffic. Area 1 includes shared tracks, which are paths that need to be shared by pedestrians and cyclists. Area 2 further includes a shared space, where no priority is given and all traffic participants need to cooperate and share the segment.

Data processing and assignment

Different sources were used to define the street network, traffic conditions, and the detailed characteristics of the cycling infrastructure in the previously described case study areas within the city of Braunschweig. For characteristics about the infrastructure, we retrieved data from Open Street Map, Open Cycle Map, "Fahrradstadtplan" (Cycle Map) of Braunschweig, and Google Maps. It is important to clarify that while the accuracy of the used open source data as well as their completeness cannot compete with data from government mapping agencies, the data still represents a solid source for deriving cycling infrastructure information in urban areas. The assignment of traffic volumes for the RG Score is currently based on static data sources such as the traffic volume map of Braunschweig. For improving the scores in the future, the accuracy of the assignment could be improved by considering expected traffic volumes on different times of the day. In addition, some individual data, particularly for width and elements alongside the infrastructure in the FMB Score, needed to be manually checked, measured, and collected from Google Earth. The current traffic volume map (City of Braunschweig, 2022b) was then used to collect data on traffic volumes and the "Unfallatlas" (Accident Map) for crash data involving cyclists from 2017 to 2019 (Statistische Ämter des Bundes und der Länder, 2022).

Given the necessary street attributes are available, both scoring methods can automatically assign ratings to the infrastructure characteristics of the street network. Importantly, before computing the ratings, certain details needed to be specified for both scores.

In the RG Score, we considered segments with more than 4000 vehicles per working day as segments with high traffic volume. This was assigned as a characteristic only for infrastructure with mixed traffic, as traffic volume plays a less critical role in cyclists' safety estimation in separated infrastructure. In addition, for the traffic signal infrastructure characteristic, we considered that the area directly impacted by a traffic signal have a similar influence on cycling behavior and safety perception as the point location. Therefore, the street network within a buffer area of 30 m radius around the location of the traffic signal was considered. This distance includes road parts that are expected to require adapting the driving behavior to the traffic signal situation (e.g., accelerating and decelerating).

For the FMB Score, only the statistically significant parameters, with a p-value less than 0.05, were considered during the rating computation. Moreover, for the infrastructure with mixed traffic, we calculated the worst-case scenario for the one-way traffic direction, considering only the contraflow coefficient when possible. The safety ratings were then computed after assigning all the required information to the street network of the case study areas.

Similarities and differences between the two scoring methods

The two proposed scoring methods are both based on studies that collected subjective safety ratings. However, although a few similarities between the methods can be observed, there are also several important differences that should be pointed out.

While the ratings for both scoring methods are based on various types of cycling infrastructures, the ratings for the RG Score are based on a more fine-grained definition of infrastructure characteristics, considering signalized intersections separately. The FMB Score, however, mainly focuses on street segments and infrastructures related to cycle tracks, cycle lanes, and streets with mixed traffic. Nevertheless, for ensuring comparability of the two scoring methods, we were able to assign most of the different infrastructures to the four common categories of infrastructure types: cycle tracks, cycle lanes, cycling on streets, and shared spaces.

Another similarity in the methodology of both scoring methods was the use of pictures that illustrate a road situation in first person view from the cyclist's perspective. In both cases these pictures have been shown to survey participants for the purpose of providing a safety rating. However, in the case of the RG Score, real photographs have been shown, while in the case of the FMB Score photorealistic illustrations of the situation were used instead. Another difference between the scoring methods is related to different rating scales for assessing safety. This issue, however, has been resolved by transforming the two different scales to one common scale to ensure better comparability.

For assigning information to infrastructures, similar data sources have been used such as OSM or Google Earth, while in both cases some manual data assignment was necessary.

Rating results

As the final scores from the two methods were in different scales, we transformed the values into a common one to make the results comparable. The goal was to assign the final score values to five distinct rating categories: 5 (*very safe*), 4 (*safe*), 3 (*neutral*), 2 (*unsafe*), and 1 (*very unsafe*). For the RG Score, the original scale ranging from 1 (*very positive*) to 5 (*very negative*) was reversed. For the FMB Score, as the survey did not include a neutral category for rating, we related the original scale of 0 to 3 to the extreme categories while excluding the neutral one. In this way, the final scores 0, 1, 2, and 3 were transformed into 1, 2, 4, and 5, respectively.

Fig. 4 illustrates the final scoring results computed using the two scoring methods overlaid on the crash data. The final result for the RG Score describes the median of the results for the previously introduced safety categories. Maps of the detailed scoring results for the different categories are provided in Appendix C. In general, we can observe that both scoring methods assigned a safe score to cycling infrastructure separated from motorized traffic, such as cycle tracks and shared sidewalks, as shown in Fig. 3B. Infrastructure shared with motor traffic, e.g., in side streets, is mostly rated as unsafe by the FMB Score. For the RG Score, this type of infrastructure is mostly rated as neutral, with an exception of an unsafe rating for few streets that are mainly classified as main streets. Area 3 is perceived as safer than the other areas, since for both scoring methods, the cycling infrastructure has been rated as relatively safe. This observation is consistent with the results from the

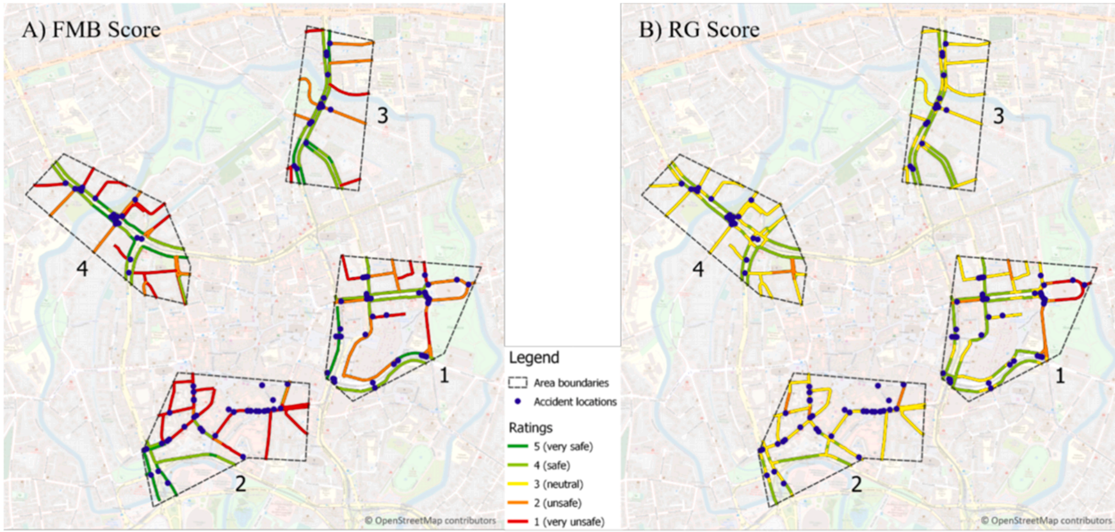


Fig. 4. Final scoring results with additional crash data overlay for the FMB-Score (A) and the RG-Score (B).

Bike Score that suggest a relatively high bikeability for area 3, as shown in Fig. 3A. For both safety scores, ratings do not seem to directly correspond to the spatial distribution of crashes with cyclists, which primarily occur at intersections. An exception seems to be a location in the north-eastern part of area 1 with a high accumulation of crashes that has been rated as unsafe by both scoring methods, as well as a street in the north-eastern part of area 2 with an unsafe rating only by FMB Score.

For the FMB Score, since it does not assign a specific score for intersections, no significant association can be inferred at locations with a higher number of crashes. For the RG Score, signalized intersections have been considered separately, but the resulting scores do not reveal notable differences in safety ratings as compared to neighboring street segments.

Validation

Ground truth

To validate the two scoring methods, they were compared with observed safety ratings collected through an online questionnaire. The structure of the questionnaire was designed similar to the one from FixMyBerlin (FixMyBerlin, 2020a). Participants were shown 20 photographs (cases) of different locations from the case study areas and were asked to rate how safe they would feel there as a cyclist. For each photograph, they could give their rating on a 5-point Likert scale ranging from 1 (*very unsafe*) to 5 (*very safe*). This scale directly corresponds to the rating categorization used for the two different scoring methods. Additionally, participants stated whether they know the locations presented as photographs in order to check for any bias due to familiarity with the locations. Fig. 5 shows an example question of the survey. The full set of photographs is provided in Appendix D.

When selecting the locations, we considered the infrastructure class, the rating results of the scoring methods as well as the crash data from the Unfallatlas (Statistische Ämter des Bundes und der Länder, 2022). We selected locations with cycle tracks (TR), cycle lanes (LN), cycling on street (ST), and shared space locations (SH) that were rated as safe, neutral, or unsafe by at least one of the scores, and where a crash occurred or not occurred. The labels of the locations are composed of the infrastructure type (TR, LN, ST, or SH) followed by a serial number. Fig. 6 shows the geographical distribution of the selected locations within the boundaries of the case study areas in Braunschweig.

The questionnaire was conducted with the tool unipark running with the software EFS survey by TIVIAN and was open for participation

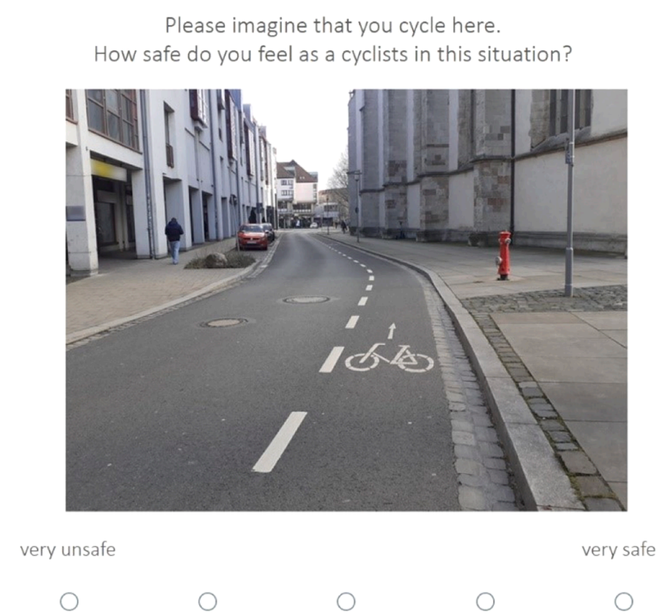


Fig. 5. Example of a question in the online questionnaire.

between 14.02.2022 and 01.03.2022. All participants were informed about the aim of the questionnaire and agreed to privacy specifications and data analysis. In total, 318 participants (153 male, 159 female, 2 diverse, 4 not specified) with an age range between 15 and 81 ($M = 23$ years, $SD = 14$ years) completed the questionnaire. Most of the participants (41.8 %) used the bicycle daily or almost daily, only a few participants (1.3 %) stated that they never cycle. The majority of the participants also commute to work by bicycle daily or almost daily (33.6 %) or several times a week (23.9 %). Another large share (19.2 %) responded that they never commute to work by bicycle.

To examine whether the familiarity with a location has an influence on the safety rating, we ran a cumulative mixed model with the familiarity as fixed factors, the participants' ID as random factor, and the survey ratings as ordinal outcome. The results of the estimated coefficients indicated that familiarity with a location does not significantly influence the safety rating. Therefore, the familiarity of the location is not considered in the further analyses.

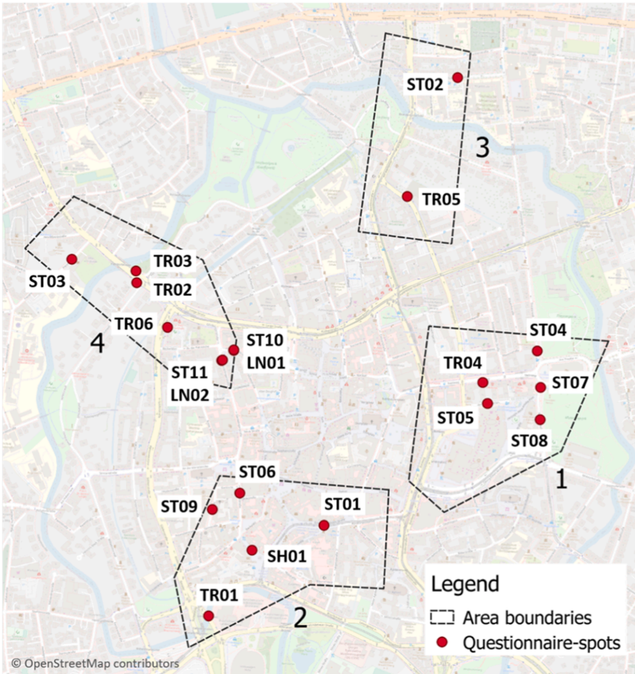


Fig. 6. The 20 locations used in the questionnaire and their distribution in the four case areas in Braunschweig.

Scoring methods evaluation

To examine which of the scores correspond the most with the questionnaire ratings, we calculated the interrater-reliability of each of the scores with the questionnaire using the weighted Cohen's kappa for ordinal rating scales. Additionally, we included the Bike Score that we initially used to identify our locations of interest. With the weighted Cohen's kappa, we consider the RG Score, the FMB Score, the Bike Score, and the Questionnaire to be independent raters that evaluate the 20 locations on an ordinal scale ranging from 1 (*very unsafe*) to 5 (*very safe*). For this, we needed to round the values of the RG Score. Because the RG Score is based on the median of values, two locations had a score of 3.5 and were rounded to a score of 4. Furthermore, the Bike Score that originally ranges from 0 to 100 was split equally into five categories to ensure that all scores range from 1 to 5. Table 2 shows the kappa statistics and the confidence intervals for each of the comparisons.

As can be seen in Table 2, the RG Score shows the highest interrater-reliability with the questionnaire ratings. According to Landis and Koch (1977), the RG Score has a moderate agreement, the FMB Score has a fair agreement, and the Bike Score has a poor agreement with the questionnaire ratings. Numerical results of the score values are provided in Appendix E.

To look closer at the agreements and disagreements of the scores, Fig. 7 shows the relation between each score and the questionnaire in scatterplots, and Fig. 8 further shows the ratings of our two scores and the questionnaire for each location as bar chart. Because of the poor agreement, the Bike Score is not considered in Fig. 8 anymore. Please consider that Fig. 7 and Fig. 8 present the not-rounded RG score values again which can be 3.5. Apart from that, the data points jitter around their category to avoid overlapping data points.

Table 2
Kappa statistics with weighted Cohen's kappa for the interrater-reliability of the Scores and the questionnaire.

		95 % CI of kappa	
	kappa	lower bound	upper bound
RG Score – Questionnaire	0.45	0.26	0.65
FMB Score – Questionnaire	0.36	0.21	0.50
Bike Score – Questionnaire	−0.07	−0.24	0.10

As can be seen in Fig. 7, the ratings of the RG Score are more centered in the middle of the graph compared to the FMB Score. In the RG Score, many locations were rated as neutral and no location was rated as very unsafe or very safe. In contrast, the FMB Score uses the full range of the scale from 1 to 5. However, the original survey that the FMB Score is based on ranges from 0 to 3 and does not include a middle category (FixMyBerlin, 2020a). This is clearly visible in Fig. 7 and might also have reduced the interrater-reliability with the questionnaire. The Bike Score has no agreement with the questionnaire (Table 2) and Fig. 7 shows that, interestingly, it has almost a reversed relation to the questionnaire. The better the bikeability, the lower the subjective safety. However, the low agreement does not surprise that much, as the Bike Score does not focus on safety aspects but also includes a destination and connectivity score and a hill score. Although the Bike Score also includes cycling facilities and a crash score, this does not increase its agreement with the questionnaire. Fig. 8 provides more detail how locations were rated on the three rating methods. Generally, most of the ratings follow a similar trend, that is, locations rated as safe or unsafe in the questionnaire are also assessed as safe or unsafe by the scores. Both scores and the questionnaire agree that the locations with a cycle track (TR) are the safest locations for cyclists, whereas locations where cyclists need to ride on the street (ST) are more unsafe. Still, at some locations the scores and the questionnaire ratings differ, which should be looked at in more detail.

First, Fig. 8 reveals that both scores underestimated some locations with cycle tracks, namely TR02, TR03, TR05, and TR06, and rated them as less safe than the participants in the questionnaire. These four cycle tracks all go along a main street where traffic volumes are higher. Both scores consider the traffic volume as an aspect that reduces subjective safety. However, studies found that the traffic volume has less impact at stronger separated cycling facilities (Zimmermann et al., 2017). As the cycle tracks at these four locations are clearly separated from motor traffic, participants are less bothered by high traffic volumes and, hence, rate the locations safer. Both scores, in contrast, consider traffic volume as influencing factor but do not consider the interaction effect of volume and separation. As a result, they identify high traffic volume at those locations and, hence, rate them as less safe.

Second, additionally to the underestimation of cycle track, the FMB Score also greatly underestimates locations with cycling on streets (ST) while they are more correctly estimated by the RG Score. That is, the FMB Score generally rates locations where cyclists need to cycle on the street as more unsafe than the RG Score and participants in the questionnaire. Interestingly, the FMB Score rates ST02 and ST03 safer as the other street-locations (Fig. 8). At these locations, the side street was turned into a bicycle boulevard with more rights for cyclists. This increases the safety ratings by the FMB Score. The participants in the questionnaire, in contrast, could not differentiate between side street and bicycle boulevard, as it was not labeled as such on the photographs, and gave similar safety ratings across all street-locations.

Third, at some locations the scores overestimated the safety compared to participants in the questionnaire. The FMB Score overestimates TR01 and LN02, while the RG Score overestimates ST01 and ST04, and both scores overestimate the Shared Space at SH01. A closer look at the locations reveals possible reasons for the deviation between scores and questionnaire.

The cycle track at TR01 and the cycle lane at LN02 are rated safer by the FMB Score than by the RG Score or the participants in the questionnaire. The cycle track and the cycle lane represent rather safe cycling facilities (Zimmermann et al., 2017; Hardinghaus & Papantoniou, 2020) which is considered by the FMB Score. However, TR01 on the one hand, runs next to a main street and approaches a large intersection. An intersection and the interaction with motor vehicles connected to an intersection decreases subjective safety (van Cauwenberg et al., 2018). This is considered by the RG Score and probably also by the participants in the questionnaire, but it is not considered by the FMB Score. LN02, on

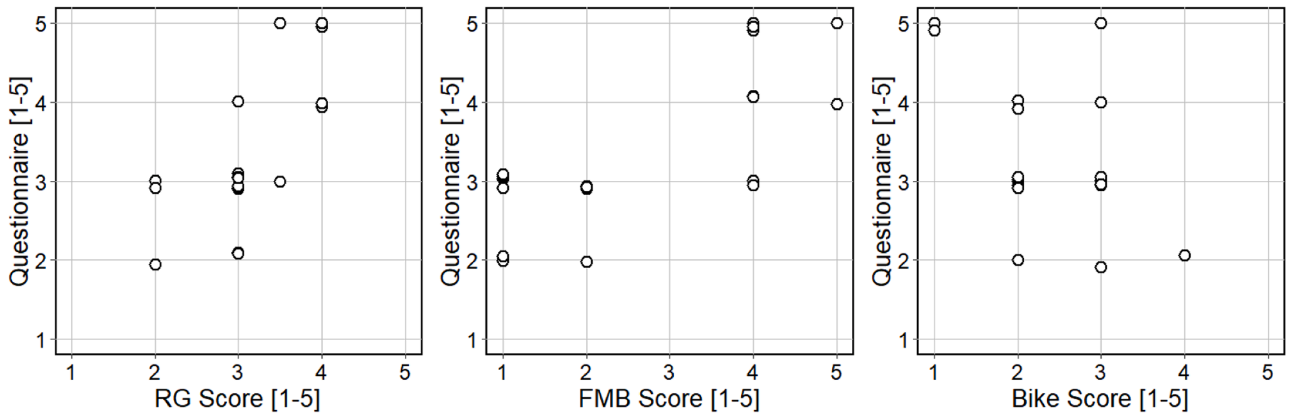


Fig. 7. Scatterplots with the relation between the RG Score (left), the FMB Score (middle), and the Bike Score (right) each with the questionnaire. Higher values refer to safer ratings. For better readability, the data point jitter around the score categories to avoid overlapping.

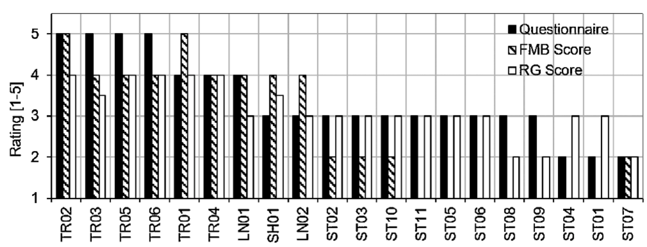


Fig. 8. Median ratings of the questionnaire and score values of RG and FMB Score for each location. Higher ratings mean safer evaluation, lower ratings mean more unsafe evaluations. Locations are sorted by their questionnaire ratings from safe to unsafe.

the other hand, runs very close to parked cars and is furthermore a one-direction street where the cyclists cycle against traffic direction. This is considered by the FMB Score, but the increase in safety due to the cycle lane might be overestimated and higher in the FMB Score than it is for cyclists in the questionnaire. They can use a cycle lane but might feel unsafe due to suddenly opened doors, cars leaving their parking lot, or motor vehicles driving on a one-direction street and not expecting cyclists coming from the other direction.

The RG Score, in contrast, overestimates the safety at ST01 and ST04. Both streets are side streets with low traffic volume and low traffic speed. However, ST01 includes tram tracks, which are a relevant hazard for cyclists in cities, as the bicycle wheels can get stuck in the tracks, or the bicycle can slip away when crossing the tracks (Leune et al., 2021). Tram tracks decrease cyclists' subjective safety (von Stülpnagel & Binnig, 2022), but are not considered by our scores.

ST04 might have similar problems as LN02 mentioned above: cyclists ride against traffic flow and close to parked cars, which might decrease subjective safety, but these aspects are not included in the RG Score which only considered the positive effect of the side street at this location.

Interestingly, the shared space SH01 is overestimated by both scores. Subjective safety might be assessed high here because of the calm traffic environment and low volume of motor traffic. Cyclists need to interact with pedestrians rather than with cars. This might be safer in the sense that crashes with pedestrians might be less dangerous and injuries less severe than with motor vehicles. However, sharing the space with pedestrians might be attention-demanding for cyclists, as pedestrians react more spontaneously and unpredictable interaction can occur. As a result, pedestrians are often evaluated negatively by cyclists (van Cauwenberg et al., 2018; Vedel et al. 2017) and a shared space might be perceived as less safe than assessed by the scores.

In summary, the RG Score seems to correspond stronger to the questionnaire ratings than the FMB Score. Apart from smaller deviations between the scores and the questionnaire, the FMB Score underestimates the safety more often and stronger than the RG Scores, especially on locations where cyclists need to cycle on the street.

Crash data

To examine whether the safety ratings of a location are connected to the occurrence of a crash at this location, we ran three separate Poisson regressions for count data with the Questionnaire, the RG Score and the FMB Score each as ordinal predictor and the number of crashes as outcome. However, the probability for a crash increase with increased bicycle volume. To control for this, we included the average bicycle volume as offset variable, following the approach by von Stülpnagel et al. (2022). The bicycle volume was derived from the dataset "Radverkehrsmengenkarte" by Grubitzsch et al. (2021). The data were log-transformed due to non-normality. To achieve comparable time periods, we only considered crashes and bicycle volumes in 2018 and 2019. A complete regression with all predictors might suffer from multicollinearity, as the predictors correlate with each other. This makes the regression coefficient difficult to interpret. Instead, we decided to run three separated regressions with one predictor and compare the Pseudo R^2 statistics.

Table 3 shows the results of the likelihood ratio test of each regression model and the Pseudo R^2 statistics. The likelihood ratio test compares each model against its respective intercept model and tests if the model with the predictor is better than the model without the predictor. As can be seen, only the questionnaire seems to be a significant predictor. However, it needs to be noted that none of the categories of the questionnaire gave significant parameter estimates. Together with the very low Pseudo R^2 statistics, this suggest that even the questionnaire

Table 3
Results of the likelihood ratio test that tests each of the regression models against the respective intercept model, together with Pseudo R^2 statistics for each model.

	Likelihood Ratio test against intercept model			Pseudo R^2		
	X^2	df	p	McFadden	Cox & Snell	Nagelkerke
Questionnaire	8.17	3	.043	0.11	0.34	0.34
RG Score	4.78	3	.188	0.07	0.21	0.22
FMB Score	2.08	3	.555	0.03	0.10	0.10

Outcome: Sum of crashes in 2018 and 2019.
Offset-Variable: Average bicycle volume in 2018 and 2019, log-transformed.

rating is not a good predictor for the number of crashes. The RG Score and the FMB Score are both even worse predictor, as the model comparison showed no difference and the Pseudo R^2 are even lower.

Fig. 9 visualize the relationship between the score values and the number of crashes in scatterplots. For the Questionnaire and the FMB Score, we can see at a least the tendency that locations with a higher number of crashes are rated safer. However, the FMB Score also rates many locations as unsafe which have only very few crashes. Together with the limitation to only 20 case locations, this might prevent a stronger relationship between rating and number of crashes. As mentioned above, the RG Score does not use the full range of the scale but rates many locations as rather unsafe or safe or neither unsafe nor safe. As a result, it is not able to distinguish very unsafe or very safe locations and this might prevent a strong relation to the number of crashes. In conclusion this means that neither the scores nor the questionnaire were able to predict the occurrence of crashes much better than chance.

Previous studies found the objective and subjective safety to go hand in hand but diverge in specific situations (von Stülpnagel et al., 2022; Winters et al., 2012). This deviation can result from cyclists underestimating the risk of crashes at certain locations (von Stülpnagel et al., 2022), or from underreporting crash statistics, as they mainly do not include near-misses or hazardous interaction without a crash (Gössling & McRae, 2022; Winters & Branion-Calles, 2017). However, as the findings in Section 5.2 already suggest, cyclists' perception of safety is a result of weighing up several aspects including the separation from motor traffic, possible interaction, or the presence of possible hazards. This is in line with the suggestion of von Stülpnagel and Binnig (2022) that designing a street segment by adding objectively safe elements might not directly lead to subjective safety. If another aspect remains unsafe in the eyes of the cyclists, an objectively safe location can still be perceived as unsafe.

In line with this literature, it is not surprising that the relationship between our scores and crash numbers is relatively low. Furthermore, the focus on only 20 locations results in a too small dataset to adequately assess a strong connection between ratings and crash numbers.

Discussion

Interpretation of the results

For most of the cases, the computed scores, and especially the FMB Score, were more conservative than the participants' ratings in the questionnaire. This raises the question of whether the scores overestimate the risk of certain infrastructural attributes and rate certain locations as less safe than they actually are, or whether cyclists potentially underestimate risk at certain locations.

The comparison of the scores and the questionnaire revealed that cyclists may consider more aspects than the scores. This includes specific infrastructural elements such as tram tracks which are simply not covered by our scores but have clear negative impact on subjective safety (von Stülpnagel & Binnig, 2022), but this also includes to allow that characteristics might be weighted differently and interact with each other. That is, a location with parked cars or a shared space might be perceived unsafe even if the cyclist can ride on a cycle lane or in low traffic volume. In contrast, a separate cycle track might be perceived as safe even if the surrounding traffic volume is high. All these interaction effects are not sufficiently covered in our scores and lead the scores to under- or overestimate the subjective safety at some locations. Still, the scores achieve a fair to moderate agreement with the questionnaire (Table 2) and locations that are rated safer or less safe by participants are also evaluated safer or less safe in the scores.

Interestingly, the Bike Score did not show any agreement to the questionnaire although it also includes cycling facilities and crash data. However, as this score does not focus on safety, it is less surprising that it is not associated with it. The question is which score or which evaluation determines at the end whether a cyclist is motivated or deterred from cycling or which route is more or less attractive to cycle. Subjective safety might be a fundamental determinant here and people will decide to use the bicycle if they feel safe or will switch to another mode of transport if they feel unsafe. Still, the Bike Score and the aspect it considers might play a role on the next level of infrastructure evaluation. That is, when safety is provided and the infrastructure should be evaluated in terms of connectivity, efficiency, or convenience. To assess subjective safety, however, the Bike Score might not be an appropriate tool.

Interestingly, none of the scores and only slightly the questionnaire were sufficiently associated with the number of crashes at the locations. Again, a trend indicate that crash-prone locations are evaluated less safe by the scores (Fig. 9), but the premise that objective and subjective safety are not highly correlated is further reinforced by our results. Countermeasures such as traffic lights or signs are often installed at locations with high traffic volume or high levels of crashes. They might increase the objective but not the subjective safety. In contrast, situations that are potentially perceived as safe, such as intersections or junctions in a calm side street, might be objectively unsafe (von Stülpnagel et al., 2022). In this way, objective and subjective safety does not necessarily go hand in hand. However, the presented study only relied on the data from the "Unfallatlas" (Statistische Ämter des Bundes und der Länder, 2022) and did not consider any other sources such as hospital reports. Methodological improvements of the scores might also improve the meaningfulness of the results, whether this leads to a strengthening or weakening of the association between scores and crash rates.

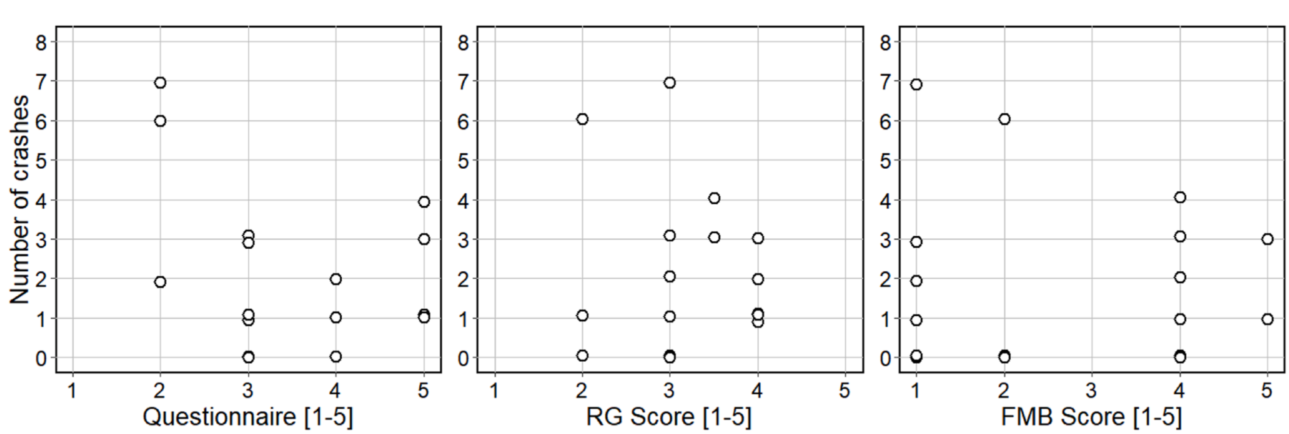


Fig. 9. Scatterplots with the relation between questionnaire and score values and number of crashes. Data points jitter to avoid overlapping.

Limitations and opportunities

Although interesting and promising results were presented, both the RG Score and the FMB Score have some limitations. As we could not identify all infrastructure elements from the open data sources, one significant drawback is the data collection. For instance, information for the width of infrastructure elements and surrounding objects in the environment was manually collected, which prolonged the work. Furthermore, both scoring methods are derived from surveys using ordinal scales with rather limited ranges, from 1 to 5 or from 0 to 3. This may reduce the variability in participant ratings, potentially constraining the depth of insights that could be derived from the data. A wider rating scale would have allowed for greater variance and more nuanced distinctions in the participants' subjective safety.

Regarding the RG Score, some of the safety categories were based on very few responses from the survey participants which reduces the representativeness of the results. A higher reliability could be achieved by improving the method with a larger sample size. In addition, the RG Score weights all assigned infrastructure characteristics equally. The current calculation method with equal weights also contributed to most of the ratings being neutral. It might be reasonable to adjust the weighting of different infrastructure characteristics based on their relevance to safety. For instance, in some cases, ratings based on street types might be more relevant for subjective safety than ratings based on the surface or traffic volume. Lastly, many street characteristics used in the RG Score were binary, e.g., high and low traffic volume, or asphalt and cobblestone. As a result, the score can only evaluate streets with these characteristics but cannot draw conclusion about characteristics that were not considered, such as medium traffic volume, or fine gravel or poor asphalt surface. In this regard, the RG Score could be extended in the future to represent a more comprehensive scoring tool.

The FMB Score, which only applies to street segments, neglects the particularities of intersections and junctions from the computation. Considering the different traffic controls and design elements to assess their weight to subjective safety is relevant since in these cases, cyclists are likely more vulnerable to crashes. Moreover, it is important to note that while the FixMyBerlin survey collected participants' ratings for street segments with tram tracks, these specific scenes were not taken into account when developing our scoring method. Including this parameter could have complemented our overall evaluation.

Ultimately, the findings confirm that the scoring methods can help to comprehend cyclists' subjective safety in different cycling facilities to a certain extent. In particular, the FMB Score has the potential to be employed as a simplistic methodology or tool to provide an initial understanding of the cyclists' subjective safety based on certain design elements. Therefore, it would be an opportunity to evaluate the current infrastructure and perform further analysis or even take measures to improve it in case of detecting (very) unsafe scores. Furthermore, they are also meaningful methods to pre-evaluate the design of cycling facilities. Despite various opportunities for optimizing the proposed scoring methods for reliably predicting cyclists' subjective safety, this study indicates that objective approaches can serve as a valuable tool to support the implementation of safer infrastructure.

Conclusion

This paper introduced the RG Score and FMB Score to reliably rate cyclists' subjective safety in any cycling facility. After applying the scores in case study areas and validating the results with an online questionnaire, we learned that it is indeed feasible to assess subjective safety by objective scores. Both scores come with benefits and limitations. The RG achieves a higher agreement to the questionnaire ratings, but often lead to a medium assessment of locations and cannot sufficiently differentiate very unsafe or very safe locations. The FMB Score, in contrast, is missing important aspects such as intersections and often underestimates safe locations. Both scores do not consider all characteristics that seem to be important for cyclists and can only provide a preliminary estimation of subjective safety in different types of cycling facilities. As numerous dimensions and indirect factors are involved when addressing subjective safety, objectively incorporating the cyclists' subjective safety in a score is difficult. Considering that feeling safe is extremely relative, depending on the situation, the score might be biased in some situations. For the most part, however, this study provides a basis for future considerations and future evaluation methods to assess the subjective safety of cyclists.

CRediT authorship contribution statement

Stefan Fuest: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization, Project administration. **Mariana Batista:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization. **Frauke Luise Berghoefer:** Conceptualization, Formal analysis, Investigation, Resources, Writing – original draft. **Morten Flesser:** Conceptualization, Writing – original draft. **Bhagya Shrithi Grandhi:** Conceptualization, Writing – original draft. **Felix Spühler:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization. **Monika Sester:** Supervision, Writing – original draft. **Mark Vollrath:** Supervision, Writing – original draft.

Declaration of Competing Interest

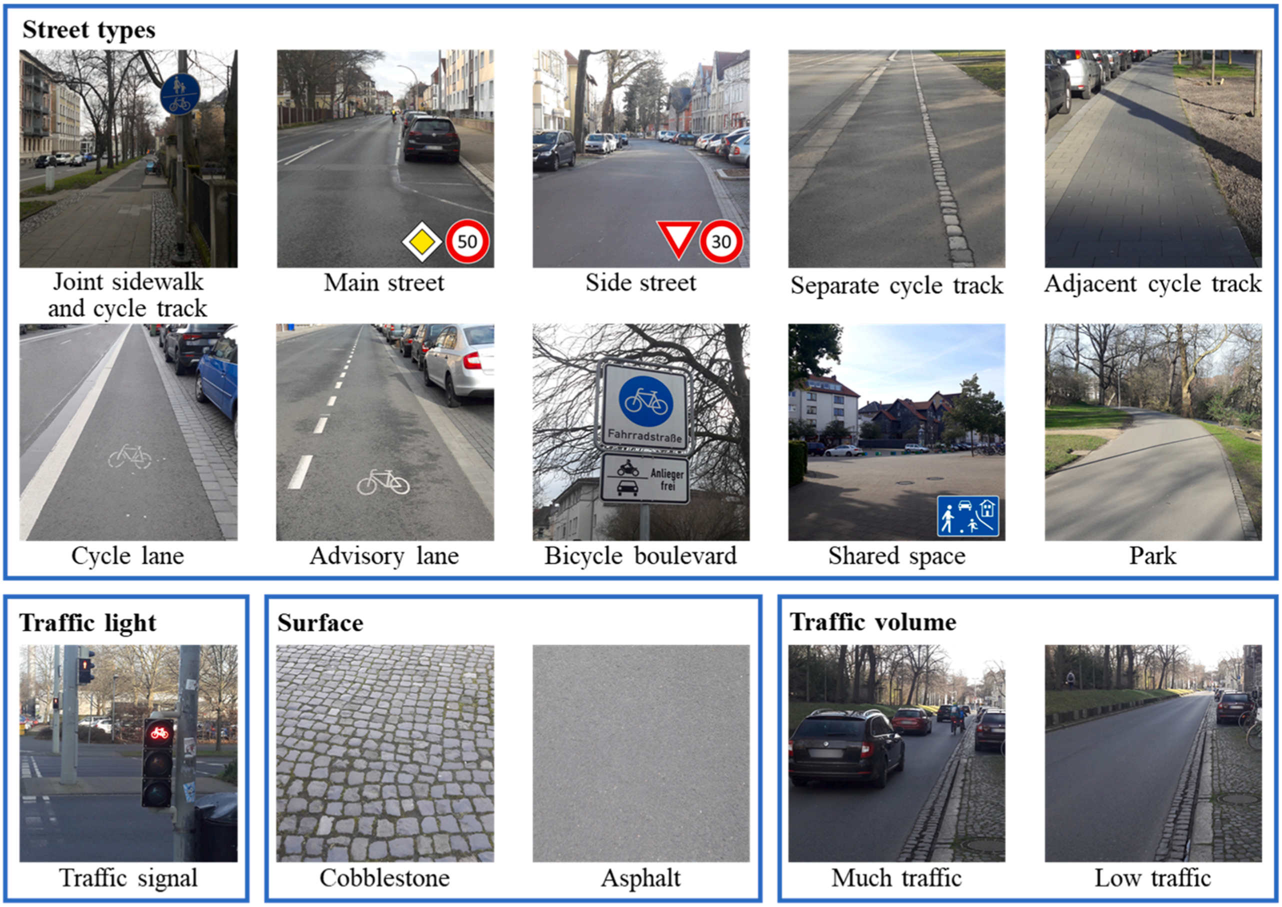
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding source

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 227198829 / GRK1931 and DAAD Graduate School Scholarship Program (GSSP).

Appendix

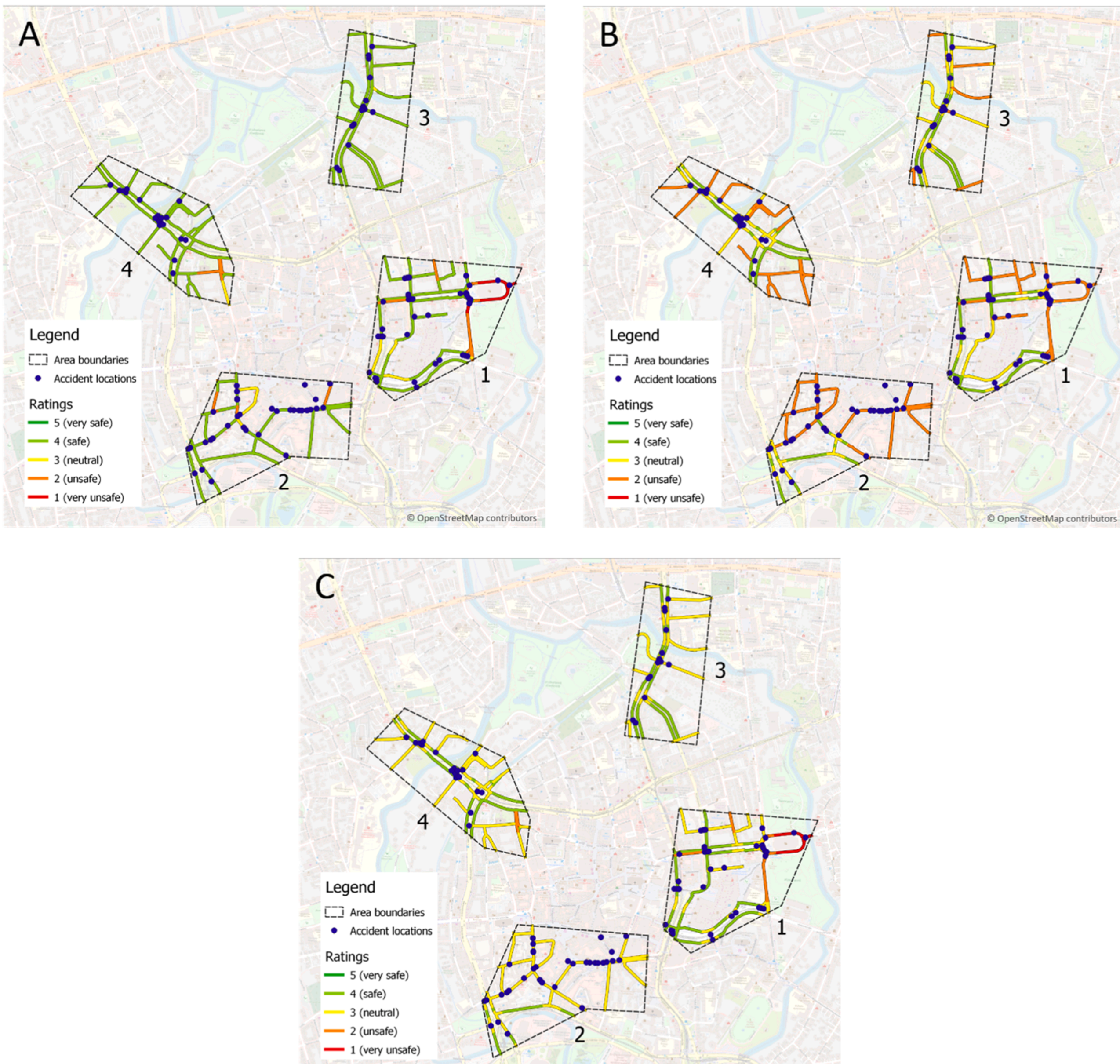
A. Infrastructure characteristics used for the safety rating of the Repertory Grid (RG) Score and derived from Berghoefer and Vollrath (2022). The various infrastructure characteristics were assigned to one of four different groups: 1) street types, 2) traffic signal, the type of 3) surface, or the 4) traffic volume.



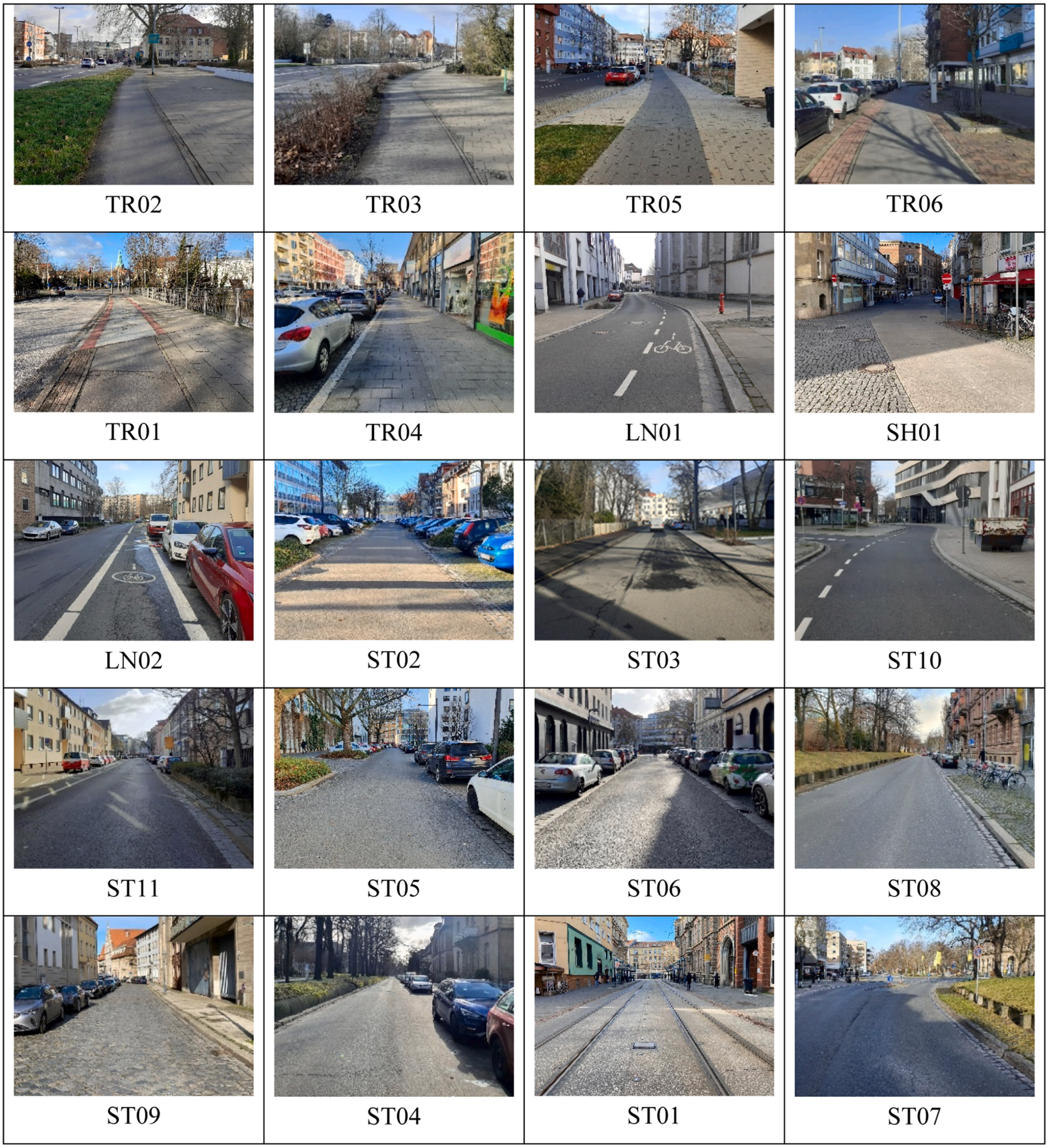
B. Excerpt of the attribute table showing the assignment of final rating scores of the RG Score. For each row (street segment), the final rating score (*Final Rating*) is calculated as the median value of the ratings of maximum four different infrastructure characteristics.

Infrastructure	Rating	Traffic signal	Rating	Surface	Rating	Traffic	Rating	Final Rating
Adjacent cycle track	2	Yes	3	Asphalt	2	–	–	2
Side street	4	Yes	3	Cobblestone	4	Low	2	3.5
Advisory lane	4	Yes	3	Asphalt	2	Much	5	3.5
Cycle lane	3	–	–	Asphalt	2	Much	5	3
Joint sidewalk and cycle track	2	–	–	Asphalt	2	–	–	2
Main street	5	–	–	Asphalt	2	Much	5	5

C. Assignments of the ratings of the RG Score (safety categories A–C) to the infrastructure classes in the case areas of Braunschweig.



D. Photographs and labels of the 20 locations used in the questionnaire ordered by their questionnaire rating from safe to unsafe (compare Fig. 8).



E. Median ratings derived for the questionnaire, the FMB Score, the RG Score and the Bike Score, as well as the number of crashes and the bicycle volume in the years 2018–2019 for the 20 different observed locations. Locations are ordered by their questionnaire rating from safe to unsafe (compare Fig. 8).

Location	Question- naire	FMB Score	RG Score	Bike Score	Number of Crashes (Sum of 2018–2019)	Bicycle Volume (Mean of 2018–2019)
TR02	5	5	4	1	3	252.5
TR03	5	4	3.5	1	4	174.5
TR05	5	4	4	3	1	49.0
TR06	5	4	4	3	1	93.0
TR01	4	5	4	2	1	45.5
TR04	4	4	4	2	2	33.5
LN01	4	4	3	3	0	7.5
SH01	3	4	3.5	2	3	87.0
LN02	3	4	3	3	0	6.5
ST02	3	2	3	3	0	56.0
ST03	3	2	3	2	0	124.5
ST10	3	2	3	3	0	160.0
ST11	3	1	3	3	0	146.0
ST05	3	1	3	2	1	82.5
ST06	3	1	3	2	3	101.5
ST08	3	1	2	3	0	133.5
ST09	3	1	2	2	1	3.0
ST04	2	1	3	4	2	195.0
ST01	2	1	3	2	7	244.5
ST07	2	2	2	3	6	119.5

References

German Cyclists Association Braunschweig (Allgemeiner Deutscher Fahrrad-Club; ADFC; 2020). *Fahrradklima test 2020: Städteranking*. ADFC supported by the Federal Ministry for Digital and Transport (BMVI). https://fahrradklima-test.adfc.de/fileadmin/BV/FKT/Download-Material/Ergebnisse_2020/Rankingliste_FKT_2020.pdf (English version at https://fahrradklima-test.adfc.de/english). Accessed July 20, 2022.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). John Wiley & Sons.

Allen-Munley, C., & Daniel, J. (2006). Urban bicycle route safety rating model application in Jersey City, New Jersey. *Journal of Transportation Engineering, 132*(6), 499–507. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:6(499)

Aziz, H. M. A., Nagle, N., Morton, A., Hilliard, M., White, D., & Stewart, R. (2018). Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: A random parameter model using New York City commuter data. *Transportation, 45*, 1207–1229. https://doi.org/10.1007/s11116-017-9760-8

Berghoefer, F. L., & Vollrath, M. (2022). Cyclists' perception of cycling infrastructure – A repertory grid approach. *Transportation Research Part F, 87*, 249–263. https://doi.org/10.1016/j.trf.2022.04.012

Bike Score ® (from Walk Score ®) https://www.walkscore.com/bike-score-methodology.shtmlAccessed July 20, 2022.

Castañon, U. N., & Ribeiro, P. J. G. (2021). Bikeability and emerging phenomena in cycling: Exploratory analysis and review. *Sustainability, 13*(4), 2394. https://doi.org/10.3390/su13042394. Article.

City of Braunschweig (2022a). Einwohnerzahlen nach Stadtbezirken (31.12.2021). https://www.braunschweig.de/politik_verwaltung/statistik/ez_stadtbezirke.php. Accessed July 20, 2022.

City of Braunschweig (2022b). Verkehrsmengenkarte (Traffic volume map). https://www.braunschweig.de/leben/stadtplan_verkehr/verkehrsplanung/verkehrsmengenkarten.php. Accessed August 15, 2022.

Daraei, S., Pelechrinis, K., & Quercia, D. (2021). A data-driven approach for assessing biking safety in cities. *EPJ Data Science, 10*(1), 1–16. https://doi.org/10.1140/epjds/s13688-021-00265-y

Desjardins, E., Apatu, E., Razavi, S. D., Higgins, C. D., Scott, D. M., & Páez, A. (2021). Going through a little bit of growing pains": A qualitative study of the factors that influence the route choice of regular bicyclists in a developing cycling city. *Transportation Research Part F, 81*, 431–444. https://doi.org/10.1016/j.trf.2021.06.005

Dill, J. (2009). Bicycling for transportation and health: The role of infrastructure. *Journal of Public Health Policy, 30*, 95–110. https://doi.org/10.1057/jphp.2008.56

Federal Agency for Cartography and Geodesy. (2022). *Digitales geländemodell gitterweite 200 m (DGM200)*. https://gdz.bkg.bund.de/index.php/default/digitales-gelandemodell-gitterweite-200-m-dgm200.html Accessed July 20, 2022.

FixMyBerlin. (2020a). *Research result - Berlin street check. Study on subjective safety when travelling by bicycle. Results and data from a survey of 22,000 participants*. July 6. FixMyBerlin supported by the Federal Ministry for Digital and Transport (BMVI) and by the Berlin Senate https://fixmyberlin.de/research/subjektive-sicherheit Accessed March 25, 2021.

FixMyBerlin. (2020b). *Umfragekonzept - NRVP-projekt kataster-ki (VB1907)*. FixMyBerlin supported by the Federal Ministry for Digital and Transport (BMVI). https://fixmyberlin.de/uploads/kataster-ki/Umfragekonzept_KatasterKI_Feb2020.pdf Accessed March 25, 2021.

Gössling, S., & McRae, S. (2022). Subjectively safe cycling infrastructure: New insights for urban designs. *Journal of Transport Geography, 101*, Article 103340. https://doi.org/10.1016/j.jtrangeo.2022.103340. Article.

Grubitzsch, P., Lißner, S., Huber, S., & Springer, T. (2021). *Radverkehrsmengenkarte*. Technische Universität Dresden, Professur für Rechnernetze und Professur für Verkehrsökologie. https://www.mcloud.de/web/guest/suche/-/results/suche/relevance/stadtradeln/0/detail/ECF9DF02-37DC-4268-B017-A7C2CF302006.

Hardinghaus, M., & Papantoniou, P. (2020). Evaluating cyclists' route preferences with respect to infrastructure. *Sustainability, 12*(8), 3375. https://doi.org/10.3390/su12083375, 1-18.

Hull, A., & O'Holleran, C (2014). Bicycle infrastructure: Can good design encourage cycling? *Urban, Planning and Transport Research, 2*(1), 369–406. https://doi.org/10.1080/21650020.2014.955210

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Leune, P., Steen, E., De Paepe, P., & Lyphout, C. (2021). An overview of tram tracks related cycling injuries in Ghent, Belgium. *Traffic Injury Prevention, 22*(3), 261–265. https://doi.org/10.1080/15389588.2021.1888941

McLeod, S., Babb, C., & Barlow, S. (2020). How to 'do' a bike plan: Collating best practices to synthesise a maturity model of planning for cycling. *Transportation Research Interdisciplinary Perspectives, 5*, 1–10. https://doi.org/10.1016/j.trip.2020.100130

OpenStreetMap (2022). https://www.openstreetmap.org.

Pucher, J., & Buehler, R. (2008). Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany. *Transport Reviews, 28*(4), 495–528. https://doi.org/10.1080/01441640701806612

Stadt Braunschweig. (2022). *Mobilitätsentwicklungsplan – zwischenbericht; ergebnisse der mobilitätsanalyse*. https://mep.braunschweig.de/wp-content/uploads/2022/04/MEP-Zwischenbericht.pdf.

Storme, T., Benoit, S., Van De Weghe, N., Mertens, L., Van Dyck, D., Brondeel, R., Witlox, F., Zwartjes, L., & Cardon, G. (2022). Citizen science and the potential for mobility policy – Introducing the bike barometer. *Case Studies on Transport Policy, 10* (3), 1539–1549. https://doi.org/10.1016/j.cstp.2022.05.013

Statistische Ämter des Bundes und der Länder. (2022). *Unfallatlas*. Germany: Statistische Ämter des Bundes und der Länder. https://unfallatlas.statistikportal.de/.

van Cauwenberg, J., Clarys, P., Bourdeaudhuij, I.de, Ghekiere, A., Geus, B.de, Owen, N., & Deforche, B (2018). Environmental influences on older adults' transportation cycling experiences: A study using bike-along interviews. *Landscape and Urban Planning, 169*, 37–46. https://doi.org/10.1016/j.landurbplan.2017.08.003

Vedel, S. E., Jacobsen, J. B., & Skov-Petersen, H. (2017). Bicyclists' preferences for route characteristics and crowding in Copenhagen–A choice experiment study of commuters. *Transportation Research Part A: Policy and Practice, 100*, 53–64.

von Stülpnagel, R., & Binnig, N. (2022). How safe do you feel? – A large-scale survey concerning the subjective safety associated with different kinds of cycling lanes. *Accident Analysis and Prevention, 167*, Article 106577. https://doi.org/10.1016/j.aap.2022.106577. Article.

von Stülpnagel, R., Petinaud, C., & Lißner, S. (2022). Crash risk and subjective risk perception during urban cycling: accounting for cycling volume. *Accident Analysis and Prevention, 164*, 10670. https://doi.org/10.1016/j.aap.2021.106470. Article.

(Jack) Winters, M., Babul, S., Becker, H. J. E. H, Brubacher, J. R., Chipman, M., Cripton, P., Cusimano, M. D., Friedman, S. M., Harris, M. A., Hunte, G., Monro, M., Reynolds, C. C. O., Shen, H., & Teschke, K. (2012). Safe cycling: how do risk perceptions compare with observed risk?. *Canadian Journal of Public Health = Revue Canadienne de Santé Publique, 103*(Suppl 3), 42–47. https://doi.org/10.1007/BF03403834.

Winters, M., & Branion-Calles, M. (2017). Cycling safety: Quantifying the under reporting of cycling incidents in Vancouver, British Columbia. *Journal of Transportation & Health, 7*, 48–53. https://doi.org/10.1016/j.jth.2017.02.010

Zimmermann, M., Mai, T., & Frejinger, E. (2017). Bike route choice modeling using GPS data without choice sets of paths. *Transportation Research Part C, 75*, 183–196. https://doi.org/10.1016/j.trc.2016.12.009