



# Distinguishing Sketch Map Types: A Flexible Feature-Based Classification

Jakub Krukar<sup>1</sup>(✉), Stefan Münzer<sup>2</sup>, Lucas Lörch<sup>2</sup>,  
Vanessa Joy Anacta<sup>1</sup>, Stefan Fuest<sup>1,2</sup>, and Angela Schwering<sup>1</sup>

<sup>1</sup> Institute for Geoinformatics, University of Münster,  
Heisenbergstrasse 2, 48149 Münster, Germany  
krukar@uni-muenster.de

<sup>2</sup> Psychology of Education, School of Social Sciences, University of Mannheim,  
Mannheim, Germany

**Abstract.** Sketch maps are often used as a means of assessing participants' knowledge of spatial environments. However, the evaluation of sketch maps is challenging as they differ in many aspects and can be scored on many possible criteria. In particular, the classification of sketch maps into different types can be problematic, because participants rarely follow any of the identifiable formats consistently. This paper presents a set of criteria that can be used to score a sketch map on two dimensions simultaneously: its "route-likeness" and its "survey-likeness". The scoring is based on the presence or absence of six features for route-conveying information and six features of conveying survey information. In the present study, reliability estimates and factor structure of the approach were examined with 460 sketch maps with a high variability of spatial elements included. Results show that the two dimensions are largely independent. Sketch maps are found that score high on the route dimensions but low on the survey dimension and vice versa, as well as sketch maps that score high (or low) on both dimensions. It is concluded that the proposed two-dimensional scoring is useful for analysing sketch maps, however, results will also depend on the task and instruction when assessing participants' knowledge of spatial environments.

**Keywords:** Sketch maps · Factor analysis · Route knowledge  
Survey knowledge

## 1 Introduction

Sketch maps are drawings of spatial environments, most often based on spatial memories of the drawer. Many readers might have drawn a sketch map before - for instance, when asked to describe a route in an area to their visiting friend. The task of drawing a sketch map is often used in spatial cognition studies as a measure of participant's spatial knowledge about the relevant area. While sketch maps are not exactly equivalent to the state of that knowledge, they can be interpreted to inform a number of commonly asked research questions.

Among them, special attention has been given to interpreting sketch map types. Since there are many ways to draw a single spatial environment, it is interesting to observe repeatable patterns in the way the environment can be depicted. For example, while some sketch maps remain reminiscent of classic metric maps, others contain egocentric views of encountered locations, or choose to abstract locations and their connectivity by depicting a graph. Some sketches are rich in landmarks, while others use them scarcely; some provide more than necessary information, while other only depict the necessary minimum; some sketches schematize turns, distances, and area extents, while others try to scale the real-world spatial relations accurately. Depending on the theory and the investigated research question, sketch maps can be classified into infinitely many combinations of types. This makes it difficult to decide about the number and criteria of categories when analysing sketch maps in each new study.

Another common problem is the fact that a single drawing rarely follows any of the identifiable formats consistently. Thus, classifying sketch maps into types has been typically performed in a subjective manner, by multiple raters who must agree on the controversial cases. This approach is not only time-consuming, but also likely to yield very different results in a potential replication, if only a small change occurs in the subjective interpretation of a single minor aspect of the sketches. For this reason, the current approach proposes a set of features, which can be scored based on the presence or absence of the given feature in each drawing. The proposed sketch map analysis is performed by following a check list and by interpreting the resulting sum scores. The features are related to two dimensions, and accordingly, two sum scores will result for each sketch map. One dimension (score) represents the extent to which a sketch map conveys information that is typical for visually describing a route. The other dimension (score) represents the extent to which a sketch map conveys spatial survey information. These dimensions are thought to be largely independent of each other.

In the following sections, we first review existing approaches for interpreting and classifying sketch map types in previous research. Then, we provide two check lists that can be used to score sketch maps on two dimensions: its *route-likeness* and *survey-likeness* (reflective of the route-related and survey-related information present in the sketch). The check list for route-likeness has six features, and the check list for survey-likeness has six features. We examine the approach by performing a reliability analysis and a factor analysis. Furthermore, the paper provides suggestions for researchers wishing to apply the scoring method in their analysis of sketch maps.

## 2 Previous Work

### 2.1 Existing Classification Schemes

The study of sketch maps was inspired by the exploratory analysis of “imageability” of city elements (Lynch 1960). Appleyard (1970) presented an extensive analysis of sketch maps drawn by the residents of the Venezuelan Ciudad Guayana. He distinguished eight sketch types classified along two dimensions: level of accuracy (topological - positional) and predominant element type (sequential - spatial). In his approach, each map has been manually classified into one of the eight categories. This method has been long influential

in the development of approaches to analyse sketch maps. For example, drawing sequences have been recently studied within this framework (Huynh et al. 2008).

Another classification was inspired by the developmental approach to spatial cognition. Moore (1976) suggested a three-level classification of the development of environmental representations. In order to demonstrate the differences between them, he described how sketch maps produced at each of these three levels would differ from each other. Sketch maps produced at Level I would only consist of egocentric viewpoints, reflecting a single, experience-based viewpoint on the subsets of the environment. At Level II, sketch maps would consist of clusters so that the quality of information within the clusters would be higher than the quality between the clusters. At Level III, sketch maps would consist of hierarchically organized clusters, related to each other on a consistent reference system.

This scheme inspired the analysis performed by Aginsky et al. (1997) where sixteen sketch maps were classified into one of three types: 0-D Place type, 1-D Place type, and 2-D Place type. These correspond to Moore's Levels I, II, and III. Nineteen percent of sketches analysed by the authors presented only egocentric views of – at most topologically related – locations (0-D Place type); half of all of the observed maps successfully connected visited places but often only with a straight line, while the individual locations were often enlarged and line distances were distorted (1-D Place type); 31% of maps had a consistent global structure, including accurate segment lengths, without disproportional enlargement of individual locations (2-D Place type). Further data analysis presented by the authors linked the 2-D Place type maps to a different learning strategy compared to the participants who produced 0-D or 1-D Place type sketches.

This work inspired the analysis later conducted in an indoor study described by Blajenkova et al. (2005). In this work, three types of sketch maps were named: 1-D, 2-D, and 3-D, based on the accuracy of spatial relations, as well as the type of topological features present in the sketches. A 1-D sketch map type was described as consisting of some turns present along the travelled route, by not including accurate relations between individual segments, and between the two floors on which the experiment took place. The 2-D drawings contained some relations between route segments, as well as approximately correct shape of the route, but did not differentiate between the two floors. The remaining drawings, named 3-D sketch maps, depicted the correct overall shape of the route and the two floors. Maps were classified as 3-D even if some of the turns they contained were incorrect. The authors use the classification into 1-D, 2-D, and 3-D sketch maps in order to derive the type of mental representation of the environment employed by the participants and use it for comparison with other wayfinding performance and memory measures.

A similar classification scheme was used by Zhong and Kozhevnikov (2016). In their experiment, participants were asked to sketch a route through a multi-level building. The authors classified 62 sketch maps into three categories: procedural route maps, allocentric-survey maps, and egocentric-survey maps. The first type consisted of maps including sequential information but no overall consistent layout of the environment. Those maps which included consistent global layout were classified as either allocentric-survey (if they employed a birds-eye view without clearly distinguishing between floors), or as egocentric-survey (if they included first-person views or clearly

distinguished between the floors). The authors provide an empirical evidence for differences in cognitive strategies employed by the participants drawing these three different sketch map types, in the indoor context.

In the work presented above, the analysis of sketch map types has been proven useful in answering diverse research questions. It is possible, however, that sketch maps form a hybrid between more than one of the distinguished types. There seem to be at least two reasons for this. First, in the above analyses accuracy of information is analysed jointly with the form (type of elements) with which the information is being depicted. Second, these classification schemes force each map into a single category. This often suggests the superiority of the last category even when the quality of low-level route information in some of its maps is poorer, compared to the maps belonging to simpler categories.

## 2.2 Theoretical Constraints for the Two-Dimensional Scoring Approach

First, it appears essential to distinguish the *accuracy* of the information being depicted on a sketch map from the *type* of information that is being depicted. The accuracy of sketch maps is a separate research problem, for which some analytical methods have been provided both for quantitative (Friedman and Kohler 2003) and for qualitative aspects (Schwering et al. 2014; Wang and Schwering 2015). However, it can be questioned whether accuracy of sketch maps should at all be evaluated with respect to a metric correspondence with the real-world configuration, because the cognitive function of sketches is not to reflect the spatial reality, but to externalize the mental representation of an environment for the specific goal of the drawing. This includes the cognitive processing of spatial features and results in, among other characteristics, schematizing, omitting, and distorting of spatial information (Tversky 2002). Sketch maps in particular are subject to many systematic changes (Tversky 1992; Wang and Schwering 2015). For example, communicating the presence of a long, straight stretch of the route does not necessitate drawing it to scale and with the level of the surrounding detail equivalent to other parts of the sketch. Moreover, constructing a sketch enforces self-consistency in the model as well as the use of Euclidian metrics between its sub-parts, neither of which might be necessarily present in the mental representation of the environment (Montello 1991; Kirsh 2010). Therefore, it can be misleading to evaluate sketch maps based on their metric accuracy in relation to the real-world configuration. This issue is particularly problematic when the sketches are subjectively evaluated by raters, since there are multiple possible sources of errors and the source of error cannot be objectively distinguished based on the final drawing. An error can result, for instance, from an erroneous mental representation, from erroneous inferential processing during the construction of the sketch, or from limited drawing experience and poor drawing ability (which might be particularly problematic when the restricted size of the available paper sheet requires planning the drawing in advance). For these reasons, analysing sketch maps based on the *presence or absence* of particular features representing qualitative aspects of conveying spatial information might be a good alternative to analysing accuracy - especially when sketch map *types* are the focus of the analysis.

Second, there seem to be two dimensions that repeatedly form the basis for the differentiation of *sketch map types*. One of them is characterized by the egocentric disjoint experience of individual locations. The other is the survey-oriented way of depicting distant parts of the environment on a single reference system (although important differences exist within diverse survey depictions, as demonstrated by Zhong and Kozhevnikov (2016)). This is in line with the well-established classification of spatial knowledge types where landmark and route knowledge are distinguished from survey knowledge (Siegel and White 1975; Montello 1998). Within this framework, integrating information about separately experienced places is considered the most challenging aspect of learning an environment. Importantly, however, information relevant to route knowledge and to survey knowledge is gathered simultaneously (Montello 1998; Ishikawa and Montello 2006). Therefore, the aim of this work is to propose a scoring approach for sketch maps which considers these two dimensions simultaneously and accounts for cases in which both route and survey information co-exists within a single sketch map. This would be difficult to achieve when each sketch map is forced into only one pre-defined category type.

The current paper does not recommend to always substitute the analysis of sketch maps' accuracy by the analysis of their types. Instead, it argues that the typology of sketch maps and their accuracy are two separate variables, that should be assessed with methods that do not intermix them. The aim of this work is to propose a method for measuring the typology of sketch maps that is independent of their accuracy. In any experimental dataset, it is possible that many sketch maps that are classified as a "simpler" type are highly accurate, or that sketch maps that are classified as a "richer" type contain profound and multiple accuracy errors. The choice of an appropriate analytical method (i.e., focused on accuracy, on typology, or on the combination of the two) should be dictated by the investigated research question.

### 3 Defining *Route-Likeness* and *Survey-Likeness* of Sketch Maps

This paper focuses on evaluating sketches on two dimensions simultaneously: their *route-likeness* as well as their *survey-likeness*. These can be interpreted as the extent to which indications of route knowledge and survey knowledge information are present in the drawing, taking into account the possibility that a single map can contain elements indicative of only one or both of these dimensions. The scoring is based on the presence or absence of specific features on a sketch map, and was inspired by the studies of elements used in route instructions given by people to other people (Denis 1997; Schwering et al. 2013; Schwering et al. 2017; Anacta et al. 2016). The following checklists provide descriptions of the specific features (criteria) – six for route information, and six for survey information – and can be applied by answering the question "Does this element exist in the sketch map?". The scoring results from the sum of points given for the identified criteria, with one point given for each criterion identified in the sketch map (theoretical range 0 – 6 for both dimensions). The procedure implies that only one point is scored if the criterion is present on the map, regardless of the number of its instances. At this stage of analysis, it is not considered

whether the depicted spatial features and spatial relations are accurate and whether they match the reality. Simply, a single point is given if the question concerned one specified criterion can be answered positively. With respect to the *route-likeness* dimension, the scoring procedure makes sense for sketch maps describing a path through the environment, which might depend on the participant's task and instruction. Multiple categories of landmarks, including their distinction into local and global landmarks, are defined based on the definitions presented in (Anacta et al. 2016). The way in which landmarks are depicted (Anacta et al. 2017) is not important for this assessment as long as the landmark can be uniquely identified by a potential sketch-map user unfamiliar with the area (e.g., a word label is enough if it describes a visible property of the landmark object).

### 3.1 Route-Likeness

Route-relevant information can be communicated by depictions of turns, landmarks, and side streets - information that helps a potential future user of the map to follow the visually indicated route and to make correct turning decisions. The route-likeness dimension is proposed to be reflected by the following six criteria.

- (1) Continuous route: are there no "gaps", interruptions or "holes" in the depicted path (such that segments of the path are missing, making it impossible to determine how to move from one fragment of the path to the next)? Is the route continuous and not fragmented?
- (2) Turns included: does the sketch depict clear turns indicative of approximate turning directions?
- (3) Side streets at decision points: does the sketch include some indication (at least a single line or arrow) of possible choice alternatives at junctions? A roundabout is treated as a regular junction.
- (4) Side streets outside decision points: does the sketch depict route alternatives along the straight stretches of the route, for instance indicating the number of junctions that need to be passed before turning?
- (5) Local landmarks at decision points: does the sketch depict local landmarks at junctions?
- (6) Local landmarks not at decision points, but along the route: does the sketch depict local landmarks along the route?

### 3.2 Survey-Likeness

Survey-relevant information can be communicated by depictions of various global landmarks, as well as hierarchical and configurational details and relations between elements not constituting a part of the main path. The survey-likeness dimension is proposed to be reflected by the following six criteria.

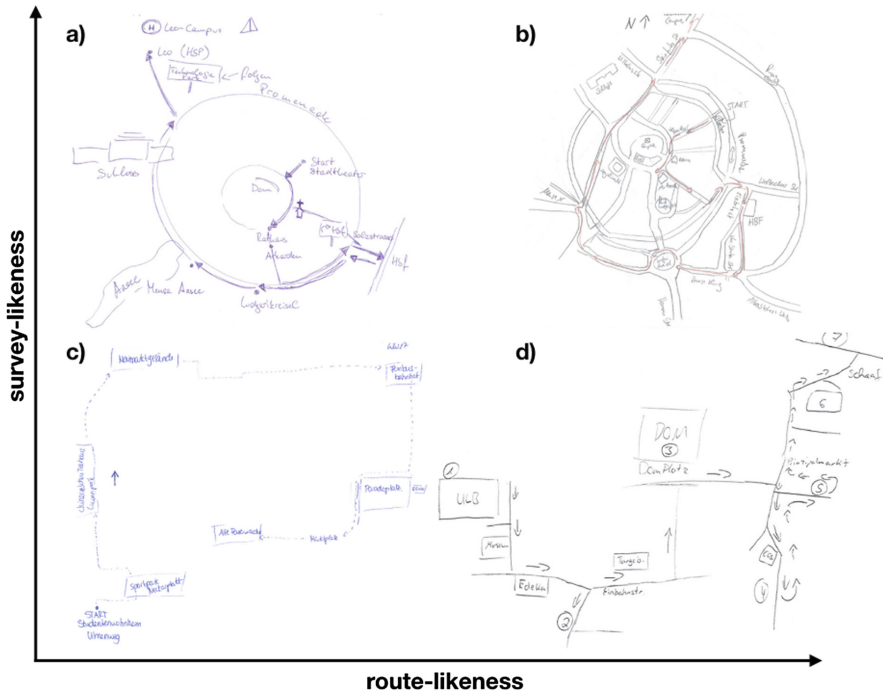
- (1) Global landmark - point: does the sketch include a point-like landmark located off-route or visible from many parts of the route? Example: a city cathedral.

- (2) Global landmark - line: does the sketch depict a line which does not constitute an integral path of the street network but provides structure to the sketch or a global spatial reference for other objects? This feature can include barriers to movement. Examples: a highway disjoint from the city streets, a river, a railroad.
- (3) Global landmark - region: does the sketch include a region, either with clearly depicted, or vague boundaries, or with a label making it a uniquely identifiable area? Examples: a zoo, a city centre.
- (4) Street network: are at least two streets connected outside the main path, so that taking an alternative route or a shortcut would be possible, at least at a short stretch of the route?
- (5) Containment hierarchy: does the sketch depict containment of one object in another object, or in a region? Examples: a tower inside a zoo, a cathedral in a marked city centre.
- (6) Spatial relation between distant objects: does the sketch depict an object which has a clear spatial relation to two other objects, which would be otherwise not directly connected to each other? Examples: A U-shaped street network with a building in the centre, where the building is clearly located between two otherwise opposite and disconnected streets.

Table 1 presents the summary of the scoring features. Figure 1 presents examples of sketch maps drawn by human participants varying on the two dimensions.

**Table 1.** Route-likeness and survey-likeness scoring features.

Dimension	Criterion
Route-likeness	r1 - Continuous route
	r2 - Turns included
	r3 - Side streets at decision points
	r4 - Side streets outside decision points
	r5 - Local landmarks at decision points
	r6 - Local landmarks outside decision points
Survey-likeness	s1 - Global landmark - point
	s2 - Global landmark - line
	s3 - Global landmark - region
	s4 - Street network
	s5 - Containment hierarchy
	s6 - Spatial relation between distant objects



**Fig. 1.** Sample sketch maps scoring: (a) low on route-likeness and high on survey-likeness; (b) high on both dimensions; (c) low on both dimensions; (d) high on route-likeness and low on survey-likeness.

## 4 Examination of the Two-Dimensional Scoring Approach

The main goal of the analysis was to examine whether the assessment of spatial features in sketch maps on the proposed twelve criteria are reflected by two proposed dimensions (factors): *route-likeness* and *survey-likeness*. Therefore, we applied an exploratory factor analysis.

### 4.1 Data

The data included binary (yes/no) scores for all twelve criteria, manually coded for 460 sketch maps. The sketch maps were sourced from three studies not reported in this paper in detail. They depicted one American, two German, and two fictional cities either already known to or learned by the participants. The instructions of all three studies asked participants to draw a map for a friend visiting the city, including a path between specified locations. Being sourced from multiple experimental conditions and from participants with large individual differences, the sketches are diverse: they employ different styles to depict the elements of the urban layout. Table 2 presents raw count data of the criteria within the analysed 460 sketch maps.



**Table 2.** Number of maps (out of 460) containing each criterion.

Route-likeness						Survey-likeness					
r1	r2	r3	r4	r5	r6	s1	s2	s3	s4	s5	s6
352	366	171	215	253	235	129	210	213	172	255	253

## 4.2 Inter-rater Reliability

Three raters (one of the authors and two student assistants) independently assessed a subset of the sketch maps. Sixty-one sketch maps were randomly chosen from the 460 sketch maps. Each rater scored the same chosen 61 sketch maps based on a document detailing the scoring procedure that included descriptions and examples of criteria presented above (Table 1; the full document is available at: <http://osf.io/3d97m>). Inter-rater reliability was assessed using a two-way random, agreement-based, average-measures intra-class correlation (Hallgren 2012), calculated separately for *route-likeness* scores and *survey-likeness* scores using the irr R package (Gamer et al. 2012). Intra-class correlations (ICCs) were in the “good” range, ICC = 0.68 for the *route-likeness* dimension and ICC = 0.61 for the *survey-likeness* dimension.

## 4.3 Results

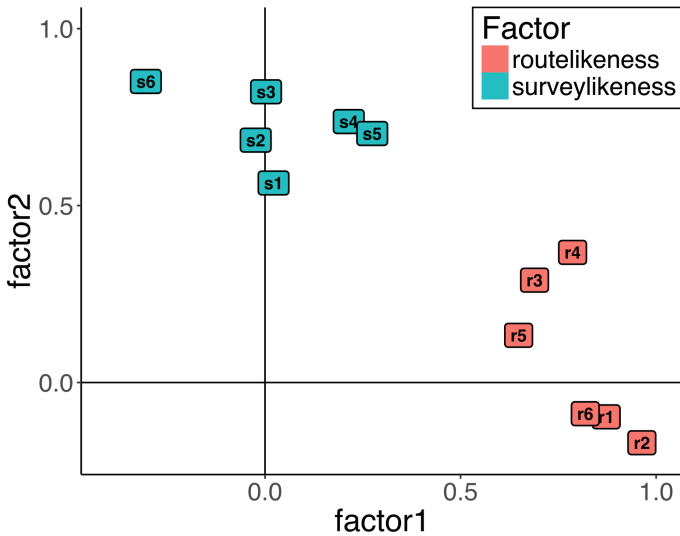
We employed exploratory factor analysis to analyse the sketch map classification. A scree plot inspection of Eigenvalues indicated that the data coming from twelve criteria might be described best using a three-factor solution. Further inspection revealed that two out of three suggested factors were sub-categories of the *survey-likeness* scale (i.e., criteria r1–r6, s1–s3, and s4–s6 loaded separate factors). On theory-motivated grounds, we performed factor analysis restricted to two factors, examining whether each criterion assumed to describe the *survey-likeness* factor indeed correlates the strongest with that factor, and not with the *route-likeness* factor. The analysis was

**Table 3.** Tetrachoric correlations between individual items.

	r1	r2	r3	r4	r5	r6	s1	s2	s3	s4	s5	s6
r1	1.00											
r2	0.93	1.00										
r3	0.52	0.58	1.00									
r4	0.68	0.75	0.69	1.00								
r5	0.39	0.51	0.43	0.50	1.00							
r6	0.50	0.64	0.51	0.61	0.69	1.00						
s1	0.03	−0.03	0.07	0.23	0.18	0.20	1.00					
s2	−0.05	−0.06	0.17	0.24	0.31	0.15	0.37	1.00				
s3	0.02	−0.04	0.24	0.39	0.32	0.03	0.56	0.53	1.00			
s4	0.36	0.30	0.47	0.55	0.14	−0.05	0.18	0.38	0.44	1.00		
s5	0.21	0.21	0.51	0.53	0.15	0.29	0.46	0.31	0.56	0.65	1.00	
s6	−0.20	−0.31	0.17	0.23	−0.02	−0.29	0.21	0.48	0.52	0.72	0.40	1.00

**Table 4.** Standardized factor loadings. The correlation between the two factors equals 0.14.

Criterion	Factor 1	Factor 2
r1	0.872	-0.097
r2	0.964	-0.170
r3	0.689	0.289
r4	0.786	0.368
r5	0.649	0.133
r6	0.819	-0.088
s1	0.022	0.564
s2	-0.024	0.685
s3	0.003	0.822
s4	0.214	0.737
s5	0.274	0.704
s6	-0.305	0.851



**Fig. 2.** The twelve analyzed criteria and their factor loadings.

performed in the psych R package (Revelle 2017), using the weighted least squares solution, the oblimin rotation, and the tetrachoric correlation method (suitable for binary data, such as the yes/no responses). This analytical method does not pre-impose any structure on the data and is thus suitable for examining whether the correlations found in the dataset are similar to those assumed on the theoretical grounds by the researchers. Raw correlations are presented in Table 3. Results of the factor analysis are presented in Table 4 and Fig. 2. Guttman’s Lambda 6 was used to calculate reliability, as it is appropriate for binary data and for the joint assessment of multiple

scales. Interpretation is similar to Cronbach's alpha with its values ranging from 0 to 1; the score of .70 is considered satisfactory for preliminary research, and scores above .90 are expected for decision-making tools (Nunnally 1978). Guttman's Lambda 6 was 0.82 for the route-likeness dimension and 0.73 for the survey-likeness dimension.

#### 4.4 Discussion

The analysis confirms that twelve previously presented criteria can be seen as representative of two factors, in a structure as suggested in Table 1. It bears noting, that the purpose of this analysis was not to find the factorial structure reducing the data best to the fewest components, but rather to verify the assumption that twelve criteria can be used as indicators of two theoretically-motivated factors.

The correlation between both factors was small (0.14) and there was some moderate correlation between individual criteria belonging to two separate factors (Table 3, e.g., criteria r4 and s4). This is not surprising, given the fact that the characteristics of sketch maps tends to correlate with individual abilities, the knowledge of the environment, and the experimental task at hand. It is therefore likely that highly-skilled individuals draw information-richer maps in general ("richer" meaning sketches with more points on the presented classification), that higher familiarity contributes to richer maps on both dimensions, and that easier tasks result in richer maps on both dimensions. Overall, however, the correlation of 0.14 between the two factors is small. It is apparently meaningful to analyse the two dimensions – *route-likeness* and *survey-likeness* – separately, despite the fact that they are not completely orthogonal. The advantage of scoring sketch maps on two largely independent dimensions lies in the opportunity for making finer distinctions between sketch maps. It is possible, for instance, to consider maps that are high on one dimension, but low on the other, high on both, or low on both dimensions; and this can be done without creating new categories or manually re-classifying maps.

The two-dimensional scoring also contributes to the solution of the problem of hybrid sketches. In our dataset of 460 sketches, 327 scored at least one point on both dimension. Thus, most of the sketch maps are hybrids: they contain information relevant to both the information concerning a specified route and to the allocentric overview of the broader environment. The presented two-dimensional scoring approach does not involve forcing each hybrid map into a single category.

Nevertheless, more discrete categorizations can be derived, if needed. For instance, it is possible to calculate the mean *route-likeness* and mean *survey-likeness* scores and classify sketch maps into four relative categories, depending on their score below or above the two means (Fig. 1 contains some examples). Performing this operation on our 460 sketches resulted in 116 maps falling into the "low - low" category, 84 sketches in the "low route - high survey" category, 95 sketches in the "high route - low survey" category, and 165 in the "high - high" category. The spread of these numbers highlights the fact that the two-dimensional scoring approach can capture the diversity of sketch maps and does not result in a disproportionately large number of "poor" maps.

The implication for broader spatial cognition studies lies in the fact that the presented classification method appreciates the diversity in drawing strategies among individuals. For instance, it could be intuitively expected that higher amount of survey

information is always associated with high amount of route detail. However, even poorly performing participants differ in their drawing strategies and it is possible that some participants aim to convey large amount of configurational information even when they do not have detailed knowledge of the route they are required to describe. This, as well as other nuanced relations, would be difficult to capture using a pre-defined set of sketch map categories, unless such a category is explicitly expected to occur among the sketches.

## 5 Limitations and Conclusion

The paper presented an approach that can be used for scoring sketch map types. Twelve criteria have been shown to load two separate factors, which correspond to the theoretically-supported concepts of *route-likeness* and *survey-likeness*. The scoring approach does not require forcing hybrid maps into a single category and makes it possible to distinguish higher and lower scores within each dimension.

The inter-rater agreement was not perfect in a situation when three raters scored the sketches based on a single document with descriptions and examples. An active communication between the raters might be necessary to ensure a shared understanding of all criteria. Since the sketches evaluated in our study came from three separate experiments, their diversity (and dissimilarity to the prototypical examples) might have been larger than it is typically the case for sketches derived from a single study.

It is important to note that the scores resulting from using the checklist are on the ordinal, and not on the continuous ratio scale. Limitations similar to likert-scale measures apply. Moreover, the distance of “1 point” should not always be interpreted in the equivalent way. If to consider three maps scoring 2, 3, and 3 points on the *survey-likeness* scale, the latter two maps are not necessarily “more survey-like” from the first map to the same extent, as their points might derive from different criteria. Likewise, two maps should not be considered disposing completely identical characteristics when they score the same number of points, as the points might be derived from distinct criteria. Researchers should consider relevant statistical tools for further analyses, depending on the case-specific application of the approach.

It also bears noting that the presented two-dimensional scoring approach does not consider the accuracy of information contained in sketch maps, but it offers the possibility of including this aspect in the analysis. This could be done, for instance, by scoring the maps twice: once for the presence/absence of the criteria listed in Table 1, and the second time for the presence of correct vs erroneous instances of each criterion. A researcher investigating the accuracy of sketch maps would then be interested in the relation between these two values. Yet another alternative is to impose a threshold value of instances of each criterion that needs to be reached before a point is awarded on the checklist. This could either be a generic number (e.g. a point is awarded only if more than two *landmarks at decision points* are present on the sketch map), or it could be linked to a particular location of interest (e.g. a point is awarded only if a *landmark at decision point* other than the cathedral is present on the sketch map). The scoring can be also filtered by the informational value of the elements included in the sketch maps: for example, landmarks that are not visible from the route can be ignored in the scoring

process. The presented scoring approach is flexible enough to support multiple application scenarios, without the need for deriving new category sets for each experimental dataset.

## References

- Aginsky, V., et al.: Two strategies for learning a route in a driving simulator. *J. Environ. Psychol.* **17**(4), 317–331 (1997)
- Anacta, V.J.A., Humayun, M.I., Schwering, A., Krukar, J.: Investigating representations of places with unclear spatial extent in sketch maps. In: Bregt, A., Sarjakoski, T., van Lammeren, R., Rip, F. (eds.) *GIScience 2017. LNGC*, pp. 3–17. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-56759-4\\_1](https://doi.org/10.1007/978-3-319-56759-4_1)
- Anacta, V.J.A., et al.: Orientation information in wayfinding instructions: evidences from human verbal and visual instructions. *GeoJournal* **82**(3), 567–583 (2016)
- Appleyard, D.: Styles and methods of structuring a city. *Environ. Behav.* **2**(1), 100–117 (1970)
- Blajenkova, O., Motes, M.A., Kozhevnikov, M.: Individual differences in the representations of novel environments. *J. Environ. Psychol.* **25**(1), 97–109 (2005)
- Denis, M.: The description of routes: a cognitive approach to the production of spatial discourse. *Cah. Psychol. Cogn.* **16**(4), 409–458 (1997)
- Friedman, A., Kohler, B.: Bidimensional regression: assessing the configural similarity and accuracy of cognitive maps and other two-dimensional data sets. *Psychol. Methods* **8**(4), 468 (2003)
- Gamer, M., Lemon, J., Singh, I.F.P.: irr: Various Coefficients of Interrater Reliability and Agreement (2012). <https://cran.r-project.org/package=irr>
- Hallgren, K.A.: Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor. Quant. Methods Psychol.* **8**(1), 23–34 (2012)
- Huynh, N.T., et al.: Interpreting urban space through cognitive map sketching and sequence analysis. *Can. Geogr./Le Geogr. Can.* **52**(2), 222–240 (2008)
- Ishikawa, T., Montello, D.R.: Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. *Cogn. Psychol.* **52**(2), 93–129 (2006)
- Kirsh, D.: Thinking with external representations. *AI Soc.* **25**(4), 441–454 (2010)
- Lynch, K.: *The Image of the City*. MIT Press, London (1960)
- Montello, D.R.: A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In: Egenhofer, M.J. (ed.) *Spatial and Temporal Reasoning in Geographic Information Systems*, pp. 143–154. Oxford University Press, New York (1998)
- Montello, D.R.: The measurement of cognitive distance: methods and construct validity. *J. Environ. Psychol.* **11**(2), 101–122 (1991)
- Moore, G.T.: Theory and research on the development of environmental knowing. In: Moore, G.T., Golledge, R.G. (eds.) *Environmental Knowing: Theories, Research, and Methods*. Stroudsburg, pp. 138–164. Dowden, Hutchinson & Ross, PA (1976)
- Nunnally, J.C.: *Psychometric Theory*, 2nd edn. McGraw-Hill, New York (1978)
- Revelle, W.: *psych: Procedures for Psychological, Psychometric, and Personality Research* (2017). <https://cran.r-project.org/package=psych>
- Schwering, A., et al.: SketchMapia: qualitative representations for the alignment of sketch and metric maps. *Spat. Cogn. Comput.* **14**(3), 220–254 (2014)
- Schwering, A., et al.: Wayfinding through orientation. *Spat. Cogn. Comput.* **17**(4), 273–303 (2017)

- Schwering, A., Li, R., Anacta, V.J.A.: Orientation information in different forms of route instructions. In: Short Paper Proceedings of the 16th AGILE Conference on Geographic Information Science, Leuven, Belgium (2013)
- Siegel, A.W., White, S.H.: The development of spatial representations of large-scale environments. *Adv. Child Dev. Behav.* **10**, 9–55 (1975)
- Tversky, B.: Distortions in cognitive maps. *Geoforum* **23**(2), 131–138 (1992)
- Tversky, B.: What do sketches say about thinking. In: Proceedings of AAAI Spring Symposium on Sketch Understanding, pp. 205–210 (2002)
- Wang, J., Schwering, A.: Invariant spatial information in sketch maps—a study of survey sketch maps of urban areas. *J. Spat. Inf. Sci.* **11**(11), 31–52 (2015)
- Zhong, J.Y., Kozhevnikov, M.: Relating allocentric and egocentric survey-based representations to the self-reported use of a navigation strategy of egocentric spatial updating. *J. Environ. Psychol.* **46**, 154–175 (2016)