Journal of ENVIRONMENTAL PSYCHOLOGY

METHODOLOGICAL CONVERGENCE IN COGNITIVE MAPPING RESEARCH: INVESTIGATING CONFIGURATIONAL KNOWLEDGE

Robert M. Kitchin

Department of Geography, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP, U.K.

Abstract

Thirteen tests designed to measure aspects of configurational knowledge were compared and contrasted using a repeated measures, multi-data collection and multi-data analysis approach. Respondents consisted of 279 first-year undergraduates newly resident to the study area (Swansea, Wales). Respondents completed four tests, drawn from four different groupings which were consistent in nature. Tests were varied in spatial cueing (the amount of information supplied) and location cueing (the amount of information requested). In addition, the data produced were subjected to systematic filtering (a set number of places were removed from the analysis) and data-defined filtering (guesses and unfamiliar places were removed from the analysis). It is demonstrated that tests produce differing results and that task demands explain the variances. It is recommended that multiple, mutually supportive tests should be used to assess individuals' configurational knowledge.

© 1996 Academic Press Limited

Introduction

Cognitive mapping researchers seek to understand our behaviour in the geographical environment and the underlying cognitive processes of thought that guide our actions (Kitchin, 1994*a*). To gather data for their empirical studies a number of techniques have been employed. For example, respondents have been asked to draw a sketch map of an area (Lynch, 1960), locate points on a base map (Buttenfield, 1986), estimate the distance (Golledge *et al.*, 1969) or direction (Kirasic *et al.*, 1984) between a series of locations, recognize features on aerial photographs (Stea & Blaut, 1973), find their way along a route (Passini & Proulx, 1988), verbally describe a route or an area (Vanetti & Allen, 1988), or build a model that represents an area (Hart, 1979). In addition, qualitative think-aloud protocols are increasingly being used to elicit cognitive map knowledge (Gerber & Kwan, 1994; Kitchin, 1995; Ungar *et al.*, 1995).

The aim of the present study was to explore the extent to which tests designed to measure aspects of configurational knowledge produced analogous results (convergent validity).

Configurational knowledge incorporates infor-mation concerning angles, directions, orientation, location and distance apart of places (Golledge *et al.*, 1987) so the processor has knowledge of the associations between, and the relative positions of, places (Golledge, 1992). Research from previous studies suggest that tests designed to measure similar aspects of cognitive distance (Howard *et al.*, 1973; Cadwallader, 1979; Montello, 1991) and configurational knowledge (Magana *et al.*, 1981; Bryant, 1984; Matthews, 1984) can produce varying results. However, there has been no specific study which has explored the *reasons* for these differences, which need to be identified to provide more definitive evidence to support theory testing.

Standard texts on experimental design (*cf.* Anastasi, 1990; Cronbach, 1990) suggest that differing outcomes can be the result of a number of factors. For example, the tests might be measuring dissimilar constructs, or varying task demands unduly influencing the data collected. This paper examines the role of task demands upon the results obtained from tests designed to measure aspects of configurational knowledge. Concern over the effect of task demands in cognitive mapping research has been expressed elsewhere with Spencer and Darviezeh

(1981) and Wood and Beck (1976) suggesting that some tests might be measuring little more than the respondents' ability to cope with the test set or the mode of presentation.

In this study, the task demands of each test were carefully controlled by varying the amount of spatial information supplied (spatial cueing) and the information requested (location cueing) to allow an investigation of their effects. For example, on some tests spatial cueing was low with no spatial information provided. On other tests, spatial cueing was high, with respondents being provided with spatial information such as a base map upon which they located places. Similarly, on some tests location cueing was open, allowing respondents to choose which places to locate. Other tests were 'location cued' requiring respondents to locate designated places. Further, when these 'location cued' tests were analysed a series of data-defined and systematic filters were used to determine the specific effects of location cueing. The data-defined filters removed from the analysis data that were guesses or unfamiliar, whereas the systematic filters removed a set number of designated places.

Methodology

Respondents

The respondents used in the investigation consisted of 279 first-year geography undergraduates resident at the University of Wales Swansea. All had been resident in Swansea for approximately one term (9–10 weeks) and were aged between 18 and 29 years old. Respondents were not briefed about the nature of the tests or taught concepts relating to cognitive mapping. All had similar geographical training, access to maps and geographical details of the Swansea area. As part of their first-year practical packs each had received an Ordnance Survey Landranger (1:50,000) map of the Swansea and Gower area. The practicals for the first term had taught them how to understand maps, provided them with map design skills and given them a broad knowledge of the South Wales area; practicals tended to be at the West Glamorgan (county) and South Wales scale rather than the Swansea (city) scale. All had been on a department geo-historical tour of the city and its surrounding area in the second week of arrival. It is noted that these respondents had access to information and skills not normally available to most students.

Study area

Swansea is a coastal city compromising of approximately 168,000 inhabitants containing all the facilities of any large regional city. The city can be divided into four sections; a highland central section, east and west axes and a centre where these three regions meet (Figure 1). The east axis occupies the Swansea Valley and corresponds to the River Tawe. The west axis follows the curve of the bay. Although the university and its accommodation is exclusively in the west of the city the study area covered the whole of the city. This allowed the study of respondents' configurational knowledge of areas that were unfamiliar or visited infrequently.

Choice of tests

An extensive search of the literature (Kitchin, 1995) led to the identification of four different groupings of tests which all measure aspects of configurational knowledge. The 13 tests chosen for comparison were drawn from all four groupings and were selected to provide a diverse range of task demands. Test choice was also partly based upon the estimated time taken to complete each test.

Graphic methods. Graphic methods are all variations upon sketch mapping. Kitchin (1995) identifies five basic sketch map variations. The *basic* sketch map technique is designed to obtain from the sketch mapper a freely drawn and solicited sketch map that has been *minimally* defined by the researcher (e.g. Jacobson, 1992). The respondent is given a blank piece of paper and asked to map a given environment. The *normal* sketch mapping technique imposes more constraints on the respondent than the basic approach. The researcher is often interested in more specific features and will word the instructions appropriately to obtain the required data (e.g. Saarinen *et al.*, 1988). In *cued* sketch mapping data collection, the respondent is given a portion of the map and asked to complete specific features (e.g. Pearce, 1981). The *longitudinal* sketch map technique allows the researcher to study how the sketch map evolves. The instruction set is similar to the normal procedure but it requires the respondent to provide the sketch map on layers of carbon or tracing paper. After certain time periods the sheets of paper are turned over and the respondent continues to draw (e.g. Humphreys, 1990). Wood and Beck (1976) and Beck and Wood (1976*a, b*) have argued that teaching respondents a sketch map *language* produces maps that are not
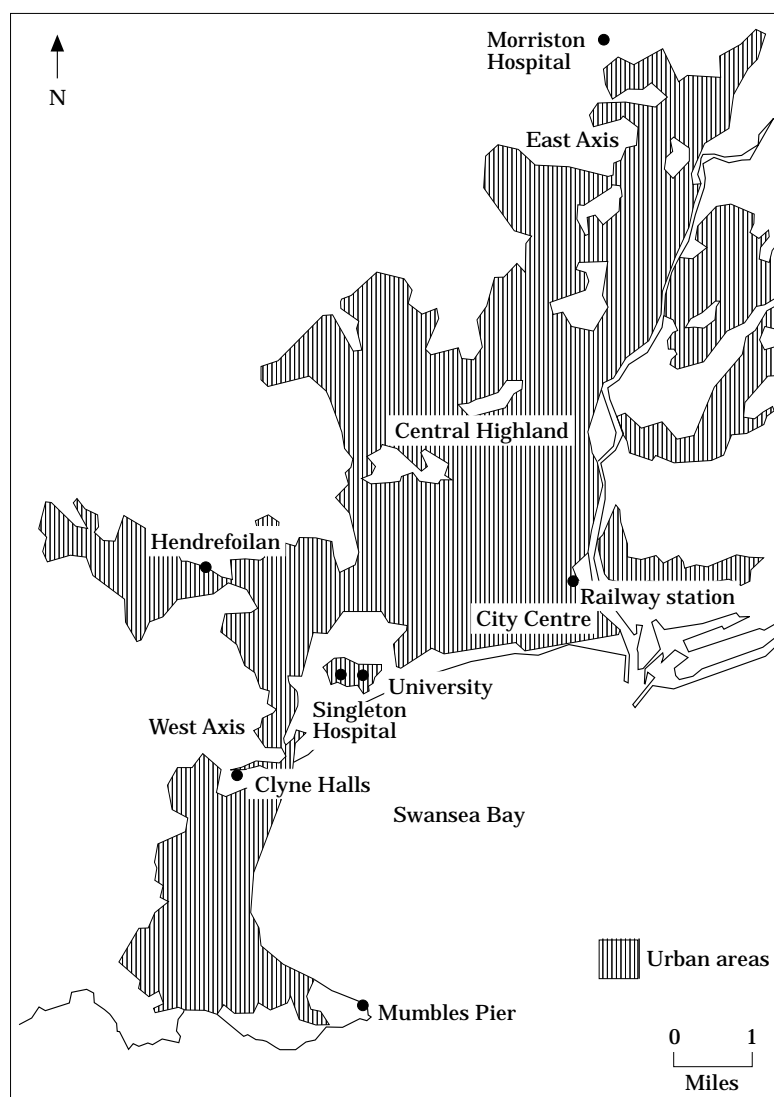
FIGURE 1. The study area (Swansea, Wales).

compromised by graphic ability or a lack of mapping knowledge. They developed a sketch map language called Environmental A for use by school children. Three graphic tests were used in the present study (one basic (*sk1*) and two normal (*sk2, sk3*) — Figure 2) varying in location cueing.

Partially graphic and reconstruction methods. Spatial cued response methodologies are essentially location testers. They differ from sketch mapping because they only require the placing of points. This reduces the motor skill component of drawing to a minimum and provides a structured framework for respondents' responses. There are various techniques but the basic method is that of Thorndyke and Hayes-Roth (1982) who asked respondents to place a location in relation to two points, one the starting point of a route and the other an arbitrary location. This method provides a scale and an orientation for the respondent. Buttenfield (1986) altered the methodology so that instead of performing a series of triad tests, a whole series of locations were placed in relation to the original pairing and a map outline.

The cloze procedure test is a spatial completion test. Traditionally the respondent 'fills in' the missing space, and an aspatial example would be, 'A dog barks but a cow __?'. Robinson (1974) and Boyle and Robinson (1978) have extended this exercise spatially. A base map is covered in grid, and the information contained in some of the squares is deleted. Respondents are then asked to identify particular

GRAPHIC TESTS

Sketch map test 1: Draw a map of Swansea inside the box. Include the coastline and other important features that you wish to add.

Sketch map test 2: Draw a map of Swansea inside the box. Include any places or landmarks that you think you know the location of.

Sketch map test 3 (sk3): Draw a map inside the box. On this map locate with a point the 25 places listed below (a list was provided to each respondent).

PARTIALLY GRAPHIC TESTS

Spacial cued response test 1 (scr 1): Try and add to the base map by placing a point where you think other places and landmarks are in relation to the geography department (A) and the Quandrant bus station (B).

Spacial cued response test 2 (scr 2): Try and add to the base map by placing a point where you think the 25 listed places and landmarks are in relation to the geography department (A) and the Quadrant bus station (B) (a list was provided to each respondent).

Spacial cued response test 3 (scr 3): Try and add to the base map by placing a point where you think other places and landmarks are in relation to the geography department (A) and the Quandrant bus station (B) and the coastline.

Spacial cued response test 4 (scr 4): Try and add to the base map by placing a point where you think the 25 listed places and landmarks are in relation to the geography department (A), the Quadrant bus station (B) and the coastline (a list was provided to each respondent).

Cloze procedure test 1 (cz1): Write in the blanked out boxes the number of a place or landmark that has been removed. Write the place that number represents in the spaces below.

Cloze procedure test 2 (cz2): Write in the blanked out boxes the number of a place or landmark from the fifteen listed below, that has been removed (a list was provided to each respondent).
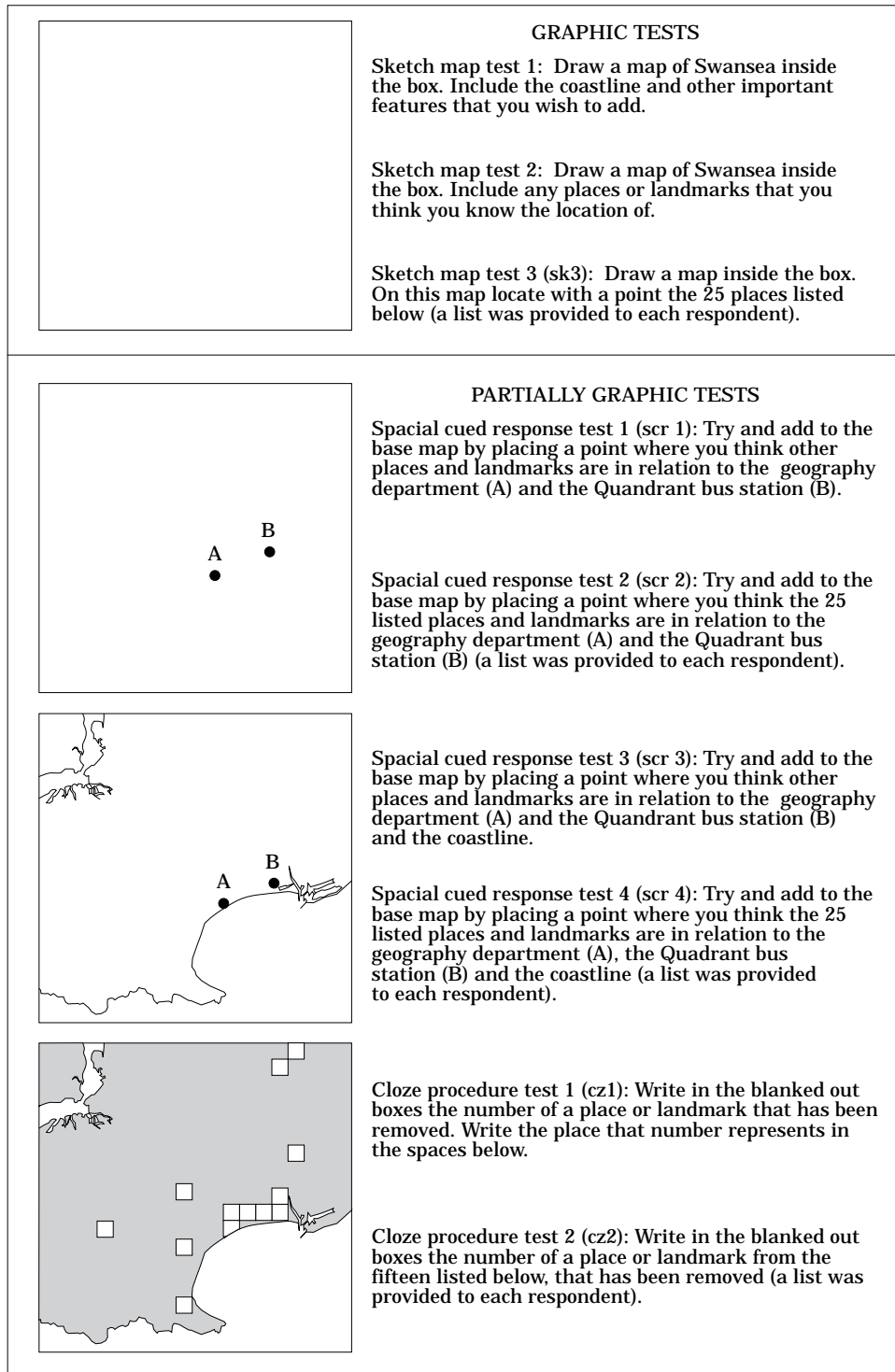
FIGURE 2.   The graphic and partially graphic tests.

elements in these blank squares with the aid of contextual information retained in the remaining open squares. Burroughs and Sadalla (1979) have used a similar technique called sentence frames. Respondents were required to complete a set of frames which took the typical format of: '__is close to __' and '__ is essentially next to __'. In the study, four spatial cued response tests are used, varying in

spatial and location cueing, and two cloze procedure tests are used which differ in location cueing alone (Figure 2).

Uni-to-multidimensional methods. The third type of data that can be used to assess configurational knowledge is that of latent data. Techniques such as multidimensional scaling and projective convergence use route or distance knowledge data to explore the latent, or inferred, structure of configurational knowledge. They do this by constructing a two-dimensional space from the one dimensional data which is provided, using a series of algorithms.

Multidimensional Scaling (MDS) is a technique that is designed to construct a 'map' showing the relationship between a number of objects, given only a matrix of 'distances' between them (Aitken *et al.*, 1989). These 'distances' can be either all metric or all ordinal. The purpose of the technique is to discover the pattern or structure in a collection of empirical data and to represent the data visually (Golledge, 1977). The algorithm minimizes the difference, or stress, between the patterns of proximities in the matrix and the space created (Montello, 1991). The resultant coordinates may exist in one of a number of dimensions and for use in cognitive mapping the results are normally constrained to two dimensions so that the latent locations in cognitive space can be compared to objective reality (Buttenfield, 1986). Many studies, such as Mackay (1976) and Magana *et al.* (1981) have used multidimensional scaling to construct a pattern of relative positioning using interpoint distance estimates. Lockman *et al.* (1981) have implemented a nonmetric equivalent. In the Lockman *et al.* (1981) study respondents were presented with the names of three locations and asked to determine which two places were furthest apart and which two were closest together. Respondents completed a series of these questions and the results were nonmetrically multidimensionally scaled to produce a two-dimensional map.

Whereas the MDS method constructs a configuration from a matrix of distances the projective convergence (or resection) method uses direction estimates to work out the coordinates of locations. Typically, respondents estimate the distance and direction to unseen places from three or more locations. The resulting vectors can be drawn and where the lines end a triangle of error can be drawn whose mean centre is taken as the cognitive location of a place. Hardwick *et al.* (1976) originally developed the method in a study where respondents first familiarized themselves with four locations

within a library. Screens were then erected and respondents asked to estimate the direction by pointing a sighting tube in the direction of the four unseen locations. By calculating where the lines intersected the triangle of error could be found and a cognitive location could be calculated. Kirasic *et al.* (1981) first used the distance/direction method to study 48 students' memory of locations on a university campus, using a direct magnitude method for eliciting distances. In a second experiment (Kirasic *et al.*, 1984), they devised a method whereby distance and direction were recorded simultaneously with respondents drawing a line which represented both. In the present study data were collected for analysis using metric (*mds*) and nonmetric (*nmds*) multidimensional scaling and projective convergence (*proj*) (Figure 3).

Recognition methods. Recognition methods collect configurational knowledge data by providing the respondents with a representation of an environment and asking them to identify features and configurations correctly. For example, Evans *et al.* (1980) asked respondents to identify four out of eight floor plans they had just walked through. Evans and Pedzek (1980) gave respondents a set of triad configurations, half of which had the correct configuration and half an incorrect configuration. These configurations were either nonrotated or rotated by 60, 120 or 180°. Respondents were shown the triads one at time and asked to say which the configuration had the places corrected located relative to each other, despite the rotation, and the reaction times were noted. The recognition test used in the study (orientation specification (*os*)) required respondents to identify which configuration out of a possible eight had the correct orientation and relationships between the places shown (Figure 3).

Research design

In order to compare validly how respondents performed on each test a repeated measures, multi-data collection, multi-analysis strategy was adopted. Initially, 177 respondents completed four of the tests which were administered independently at 1-week intervals. The tests were separated by a week mainly for pragmatic reasons — they completed the tests as part of their geography practical classes — but it also allowed a sufficient time gap to allow their previous answers to be 'forgotten'. To allow comparison across the four test groupings the respondents were split into six groups. Each group completed the same set of four tests (Table 1). Each
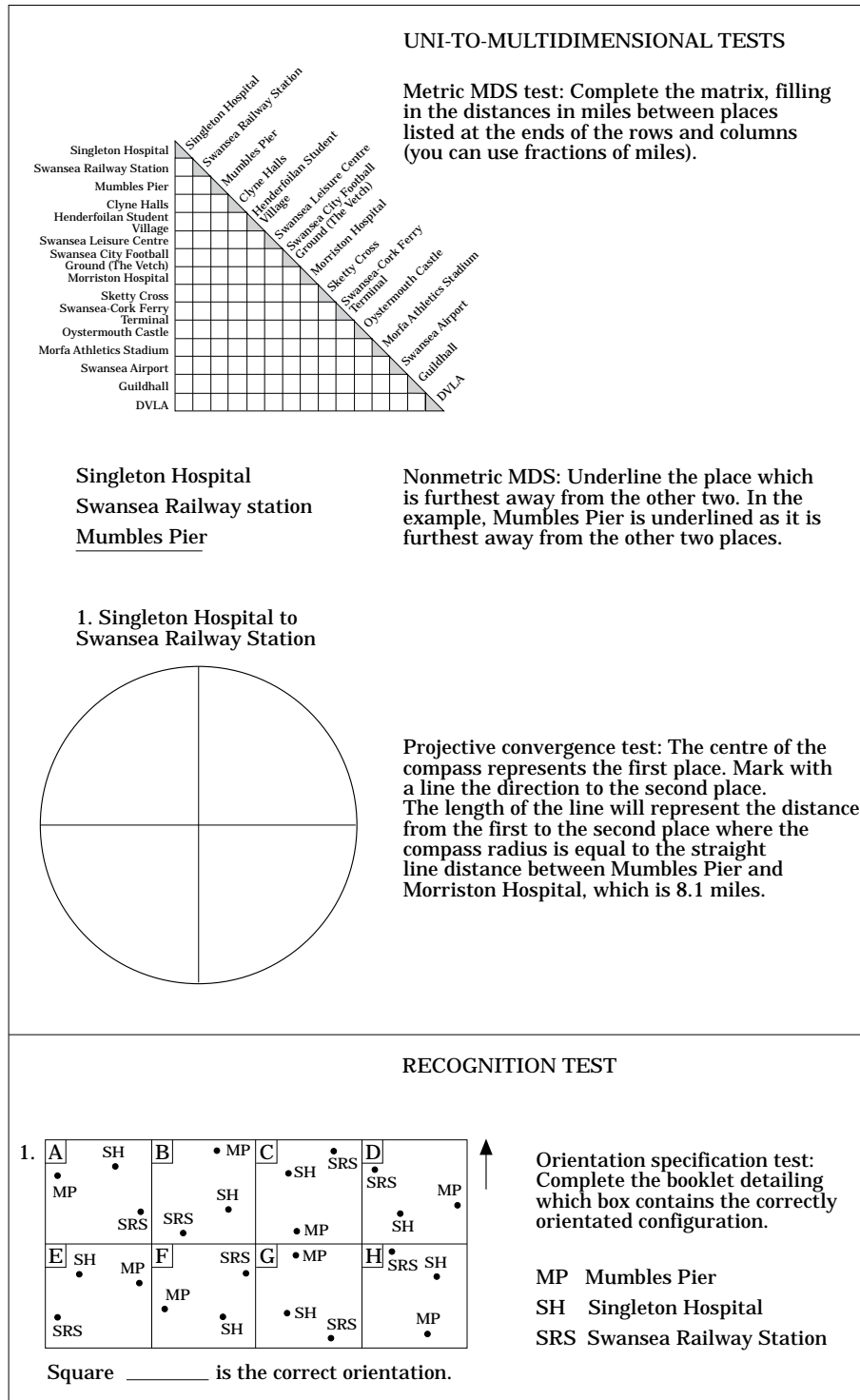
UNI-TO-MULTIDIMENSIONAL TESTS

Metric MDS test: Complete the matrix, filling
in the distances in miles between places
listed at the ends of the rows and columns
(you can use fractions of miles).

Singleton Hospital

Swansea Railway station

<u>Mumbles Pier</u>

Nonmetric MDS: Underline the place which
is furthest away from the other two. In the
example, Mumbles Pier is underlined as it is
furthest away from the other two places.

1. Singleton Hospital to
Swansea Railway Station

Projective convergence test: The centre of the
compass represents the first place. Mark with
a line the direction to the second place.
The length of the line will represent the distance
from the first to the second place where the
compass radius is equal to the straight
line distance between Mumbles Pier and
Morriston Hospital, which is 8.1 miles.

RECOGNITION TEST

1.

Orientation specification test:
Complete the booklet detailing
which box contains the correctly
orientated configuration.

MP   Mumbles Pier

SH   Singleton Hospital

SRS Swansea Railway Station

Square _____ is the correct orientation.

FIGURE 3.   The uni-to-multidimensional and recognition tests.

test was taken from a different test grouping. The
respondents in each group completed the tests in an
order which tried to minimize the transference of
knowledge from one test to another. For example,
the sketch map test 3 was completed before the spa-
tial cued response test 4 which gave respondents
the opportunity to study the coastline in detail. It
was noted that the respondents could look at maps

and actively learn between the tests. A second study collected additional data for the partially graphic exercises using a further 102 undergraduates. These data were collected for two purposes. First, the original sample sizes were small and second (and partially as a result of the first) to examine the reliability of the tests. Both of these factors allow the conclusions from the original study to be validated. To allow comparison, the data were collected exactly 1 year from the first study so that the new respondents had been present in Swansea for approximately the same length of time. Table 2 details the number of respondents executing each test.

Use of analysis framework

On the location cued tests the results were also computed using a series of analysis frameworks. These frameworks either systematically removed data from the analysis or used a data-defined filter to remove unwanted data, such as guesses, to allow the effect of such data removal to be examined. The three systematic analysis frameworks included all 25 places to be located (*25*); the first 15 places requested to remain in the analysis (*15*) and the first six places requested to remain in the analysis

(*6*). The three data-defined frameworks included those places that were guessed to be removed from the analysis (*guess*), those places that achieved a familiarity rating less than 2 (unknown and low familiar places) to be removed from the analysis (*fam<2*), and those places that achieved a familiarity rating less than 4 (low to medium familiar places) to be removed from the analysis (*fam<4*). Familiarity data for each place were collected using rating scales. The scales ranged between 0 and 5, with 0 representing completely unknown and 5 representing a highly familiar location. Familiarity was determined by asking respondents how much experience of visiting or passing they had with each of the locations used in all the tests. On analysis framework *6*, the six places used were known from an earlier spatial familiarity study (Kitchin, 1994*b*) to be familiar to the group as a whole. This framework was used to allow comparisons with the uni-to-multidimensional tests, where, because of the time-consuming nature of the tests, only six locations were used.

Analysis of tests

Because of the diversity of the data collected a number of different methods of analysis were employed,

TABLE 1
The six groups and the tests they completed

Group	Configurational test			
	Partially-Graphic	Graphic	Uni-Multidimensional	Recognition
1	Spatial cued response 1	Sketch map 3	Nonmetric MDS	Orientation specification
2	Spatial cued response 2	Sketch map 1	Projective convergence	Orientation specification
3	Spatial cued response 3	Sketch map 2	Projective convergence	Orientation specification
4	Spatial cued response 4	Sketch map 3	Metric MDS	Orientation specification
5	Cloze procedure 1	Sketch map 2	Nonmetric MDS	Orientation specification
6	Cloze procedure 2	Sketch map 1	Metric MDS	Orientation specification

TABLE 2
The number of respondents completing each test

Partial Graphic		Graphic		Uni-Multi		Recognition	
test	*n*	test	*n*	test	*n*	test	*n*
Scr 1	19 (18)			Projective convergence	41		
		Sketch map 1	47				
Scr 2	19 (15)						
Scr 3	14 (19)						
		Sketch map 2	31	Metric MDS	53	Orientation specification	109
Scr 4	33 (14)						
Cz 1	18 (22)						
		Sketch map 3	49	Nonmetric MDS	38		
Cz 2	26 (14)						

(*n*)=number from 2nd study; Scr=spatial cued response; Cz=cloze procedure.

although wherever possible, consistency was attempted with bidimensional regression being most commonly used. Data from sketch map test 3, all the spatial cued response tests, the metric and nonmetric multidimensional scaling tests, and the projective convergence test were all analysed using bidimensional regression. Bidimensional regression measures the association between configurations (Tobler, 1965) and postulates a regression-like relationship, that is basically an extension of ordinary product–moment (Pearsonian) correlation and ordinary least squares regression procedures, between two sets of coordinates. This regression technique is sensitive to rotations, translations, and changes of scale, and calculates how large these are (Tobler, 1976). In the two-dimensional situation the parameters a and b of the standard regression equation become:

$$a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \qquad b = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

This translates to bidimensional regression:

$$\begin{pmatrix} u_j \\ v_j \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x_j \\ y_j \end{pmatrix} + \begin{pmatrix} e_1 \\ f_2 \end{pmatrix}$$

where e_j and f_j are the errors.

The parameters a_1 and a_2 are analogous to the intercept term and perform the translation. The scaling and the rotation are accomplished by the matrix of b_{ij} values (analogous to the slope coefficient in ordinary linear regression). A rigid euclidean rotation is maintained by constraining b_{12} to equal $-b_{21}$, whilst constraining $b_{22}=b_{11}$ ensures the scale on both the axes is adjusted by the same amount and thus the regression grid remains equilateral (Murphy, 1978).

A number of results variables are produced: r^2 represents the goodness-of-fit between the two sets of coordinates; scale is an index that measures the scale change needed to produce the best fit with a value less than one indicating that u,v (cognitive) space needs to be contracted to fit the x,y (reality) space, and a scale value greater than one that the u,v (cognitive) space needs to be expanded; angle is the number of degrees the coordinates axes must be rotated to produce the best fit, with a positive value indicating a counterclockwise rotation, and a negative value a clockwise rotation; a_1 is the horizontal translation needed to produce a best fit, with a positive value indicating a west-to-east shift and a negative value indicating an east-to-west shift; a_2 is

the vertical translation needed to produce a best fit, with a positive value indicating a south-to-north shift and a negative value indicating a north-to-south shift (Lloyd, 1989).

Waterman and Gordan's (1984) distortion index and boxes can be calculated from the bidimensional regression results. They extended the analysis by allowing the amount of relative, absolute and total distortion to be calculated and to display this graphically. The distance between the u_j,v_j and x_j,y_j coordinates is the quantity that is minimized by the bidimensional regression transformations. They argued that this distance is the most suitable basis of comparison between different cognitive maps. The distortion distance (D) is calculated as:

$$D = \sqrt{\Sigma \left((x_j - u_j')^2 + (y_j - v_i')^2\right)}$$

where x_i and y_i are the observed x and y coordinates and u'_i and v'_i are the predicted u_i and v_i coordinates of a point.

This technique can only be used to compare different cognitive maps of the same objective map and the value does not indicate whether the distortion is large or small. The distortion index (DI) provides such a measure as it is the ratio of D to D_{max}, where D_{max} is the maximum value D can achieve which is constrained by the objective map size. D_{max} is calculated as:

$$D_{max} = \sqrt{z - (x^2 + y^2)/n}$$

where

$$x = \Sigma x_j, \quad y = \Sigma y_j, \quad z = \Sigma (x_j^2 + y_j^2).$$
$$n = \text{number of points}$$

The distortion index is calculated:

$$DI = 100D/D_{max}$$

DI, which ranges between 0 and 100, is a dimensionless value, the size of which indicates the amount of distortion regardless of the scales of the true or cognitive map. This is useful for comparing cognitive maps from the same person provided at different scales, for example. It is, in effect, a standardized measure of relative error (Lloyd, 1989).

Data from sketch map tests 1 and 2 were analysed using a map content and a map style classification. Only one rater classified the map's contents using Lynch's (1960) classification but to see if different raters would give different classifications five raters judged the sketch map's style using Pocock's

(1976) classification. The raters were all professional geographers (two lecturers and three teaching assistants) with a basic understanding of the research aims of the study.

Both cloze procedure tests 1 and 2 were analysed to produce individual and place accuracy scores. An individual accuracy score represents how well an individual did in assigning places to the boxes. A score of 100 represents all the boxes having correct locations assigned to them, and a value of zero, all the boxes having an incorrect place assigned to them. A place accuracy score represents how many of the group's individuals successfully matched that place to the correct box. A value of 100 indicates that all the respondents had assigned that place to the correct box and a value of zero that none of the respondents had assigned that place to the correct box.

The orientation specification data were analysed in two ways. First, an individual accuracy score, representing the number of configurations correctly identified, was calculated with and without guesses included. The individual accuracy scores range between 0 and 100, with a value of 0 indicating that the respondent had not correctly identified any configuration and a value of 100 indicating that all the configurations had been correctly identified. Second, a configuration accuracy score, representing the number of respondents correctly identifying a particular configuration, was calculated with and without guesses included. The configuration accuracy scores range between 0 and 100, with a value of 0 indicating that no respondent identified the correct configuration and a value of 100 indicating that all the respondents correctly identified a location.

Results

Comparing tests within test groupings

Interpreting the results from the graphic exercises.
To compare the sketch map tests 1 and 2 without bias, the two sets of maps were trimmed of their instructions, added together and randomized. This meant that the raters did not know which set of instructions the sketch mapper had received. Table 3 provides the average number of category elements for the Lynch's content classification. For example, there were on average between two and three (2·78) paths on each sketch map drawn when using the instructions provided with sketch map 1. It can be seen that both sketch map tests 1 and 2 produced

very similar patterns of contents (chi-square=4·23, $p<0·95$). The spatial products from respondents who had completed sketch map test 1 seem to contain slightly more paths but this is the result of some large anomalies caused by four respondents having large numbers of path elements in their maps. The edges and district classes differ little with very similar numbers of elements. Sketch map test 1 has slightly more nodes, but slightly fewer landmarks than sketch map test 2. This is unsurprising as the instructions provided with sketch map test 1 requested important places which are generally nodes, and the instructions provided with sketch map test 2 specifically asked for landmarks which would include nodes.

The results from the five raters using Pocock's (1976) classification are displayed in Table 4. Two observations can be made from this table. First, the instructions for sketch map tests 1 and 2 both produced maps which are very similar in style. These are generally spatial in nature containing landmarks, nodes and districts with few linear features. Second, raters classified the maps differently. Chi-square analysis to demonstrate this was not possible because even when the classes were combined too many cells contained estimated values less than 5. However, what is clear from visual analysis is that raters 1 and 2 gave very similar ratings, but raters 3, 4 and 5 all differed from each other and the first two raters.

Further analysis of the ratings revealed that very few of the maps (17% for both sketch map tests 1 and 2) received the same rating by all five raters. However, around 30 per cent of the maps were given the same classification by four raters. A further 28 per cent of maps drawn by respondents receiving the instructions for sketch map test 1 were given the same rating by three raters. For sketch map test 2, however, raters differed more in their ratings, with 19 per cent of the maps only having two raters in agreement and 10 per cent of the maps having a different rating from all five

TABLE 3
The average number in each class for the contents classification

Category	Sketch map 1	Sketch map 2
Paths	2·78 (128)	2·16 (67)
Edges	1·52 (70)	1·51 (47)
Districts	3·84 (177)	4·00 (124)
Nodes	4·91 (226)	4·58 (142)
Landmarks	5·21 (240)	5·74 (178)

(n)=actual number in each class.

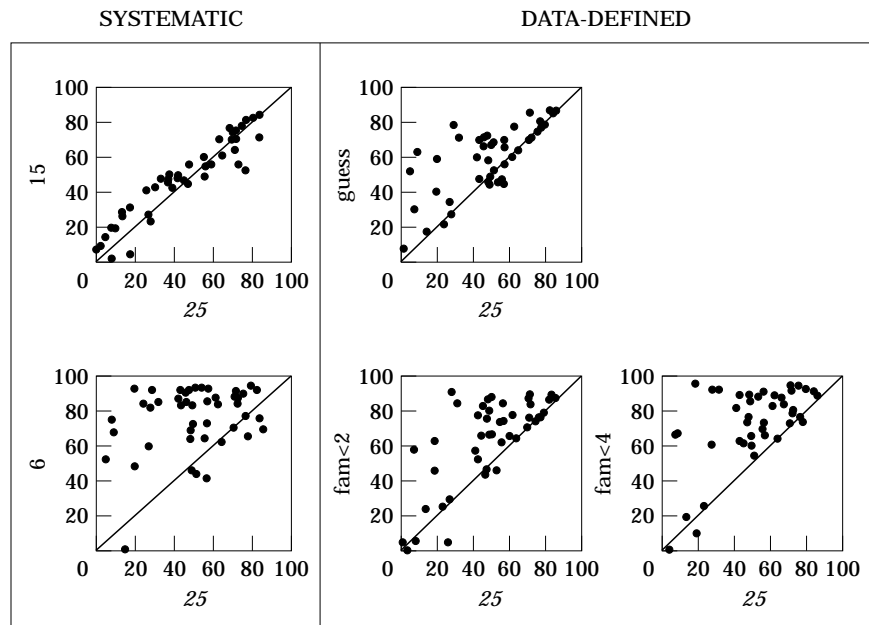SYSTEMATIC   DATA-DEFINED

FIGURE 4.   Using the r^2 values to compare the effect of the analysis frameworks upon the sketch map 3 data. Each point represents the r^2 value for an individual on each of the two analysis frameworks.

raters. It seems that the maps from respondents receiving instructions for sketch map test 2 were harder to classify, even though they had approximately the same contents as maps from the instructions for sketch map test 1. On both tests, the maps which were assigned a low style rating (A–B, G–H) achieved rating consensus. The maps that received different ratings were those that were medium to complex in style (C–D, I–J) and whose visual description in the classification chart were similar. This highlights one of the main criticisms of sketch mapping, namely the difficulty of subjectively assigning a sketch map to a category and suggests that there needs to be great caution exercised when drawing conclusions from sketch map data.

Data retrieved from sketch map test 3 were not analysed using subjective classifications but rather analysed using bidimensional regression. Because the test was so specific in its request the results could not be classified using any of the conventional subjective classifications which categorize on the basis of content, style, development and accuracy. Accuracy could be judged subjectively, but bidimensional regression provides a precise quantifiable

TABLE 4
Comparing the number of ratings in each class for the five raters of sketch maps 1 and 2

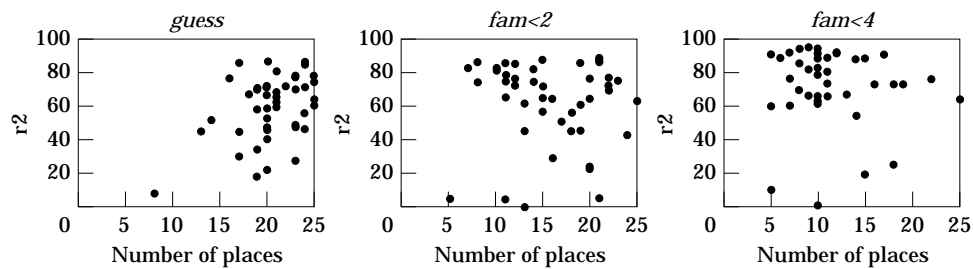| Categories		Sketch map 1 ratings					Sketch map 2 ratings				
		r1	r2	r3	r4	r5	r1	r2	r3	r4	r5
Sequential	A			2	11					6	
	B				3					2	
	C	1	1	5	5	1	1			4	
	D			1	1		1		3	2	
	E	1			1	3				2	3
	F										1
Spatial	G	2	1	5	1			2	3		
	H	3	2	1	3	5	2		5	5	
	I	32	34	25	12	20	23	25	17	6	15
	J	6	6	7	3	15	3	4	2	2	13
	K	1	2			2	1			1	
	L				6					1	

r(*n*)=rater.

FIGURE 5.   Comparing the r^2 values and the number of places remaining in the analysis.

alternative. However, Waterman and Gordan's (1984) distortion calculations could not be compared between maps as all the respondents' maps varied in scale, with no bench-mark scale or orientation having been provided.

Using a paired t-test the bidimensional regression r^2 values of analysis framework *25* were compared with the five other frameworks. Low p values (all < 0·0078) reveal that there were significant differences between r^2 values. Figure 4 reveals r^2 values are likely to increase in value, depending on the analysis framework it is calculated from. There are, however, a few anomalies where individual r^2 values decreased rather than increased. On analysis frameworks *15* and *6* these anomalies are caused by accurate data, which are compensating for the effects of large residuals, being removed from the analysis calculation, thus allowing the full 'effect' of the residuals to appear in the results. If these accurate locations are systematically removed from the analysis then their masking effect will be removed, hence the decrease in r^2 value for that framework. This 'large residual' effect might not be expected on the data-defined frameworks because these screens are designed to remove such residuals by filtering out guesses and places that are unknown. Possible reasons why a lowering of r^2 values does occur, although to a lesser extent, on the data-defined frameworks (*guess, fam<2, fam<4*), may be through distorted metacognition (over-confidence in knowing where a place is located) or genuine individual idiosyncrasies in cognitive map knowledge. The high proportion of r^2 values being equal on analysis frameworks *25* and *guess* is caused by some respondents claiming that they did not guess any of the locations. Figure 5 demonstrates that increases in r^2 value are the result of filtering out unknown knowledge, and not all the result of decreasing the number of places in the analysis. The graphs clearly show that there is no relationship between the number of places located and the r^2 value.

Interpreting the results from the partially graphic tests. Two-tailed t-tests between the r^2 values for each of the spatial cued response tests using analysis framework *25* reveal that there are significant differences between open and location cued tests (scr1 vs scr2, t=5·47, p=0·00; scr 3 vs scr 2, t=7·76, p=0·00; scr1 vs scr4, t=5·64, p=0·00; scr3 vs scr4, t=7·72, p=0·00). As might be expected, location cueing produces r^2 values significantly lower than when respondents can choose which places to locate (Figure 6). The significant difference between the two open tests is due to the amount of spatial cueing. This spatial cueing effect is not repeated when respondents are supplied with the places to locate. This is contrary to what may be anticipated as the increased spatial cueing should provide anchors to familiar locations, and those which were only partially known, and provide logical cues for others (e.g. places such as Langland Bay may be expected to be on the coast where a bay shape appears). This suggests that requested places that are difficult to locate remain difficult to locate even with the presence of a spatial cue. Thus, location cueing masks the effect of spatial cueing.

The reason why spatial cueing alters the results of open location cued tests and not the cued location cued tests (scr2(*fam<4*) vs scr4(*fam<4*), t=−1·23, p= 0·23) seems to lie in the use of the location cues. It may be the case that the location cues are producing a selection of residuals, common across the tests regardless of the spatial cues, leading to a similar pattern of r^2 values. This case is supported by the evidence that once the anomalies have been removed on the location cued tests using analysis framework *fam<4* there is a significant difference (scr2(*fam<4*) vs scr4(*fam<4*), t=−2·63, p=0·017), meaning that the spatial cueing did not have an effect when familiar data were used alone. This reveals two important facts. First, familiarity is a significant variable in determining configurational knowledge and second, that spatial cued response
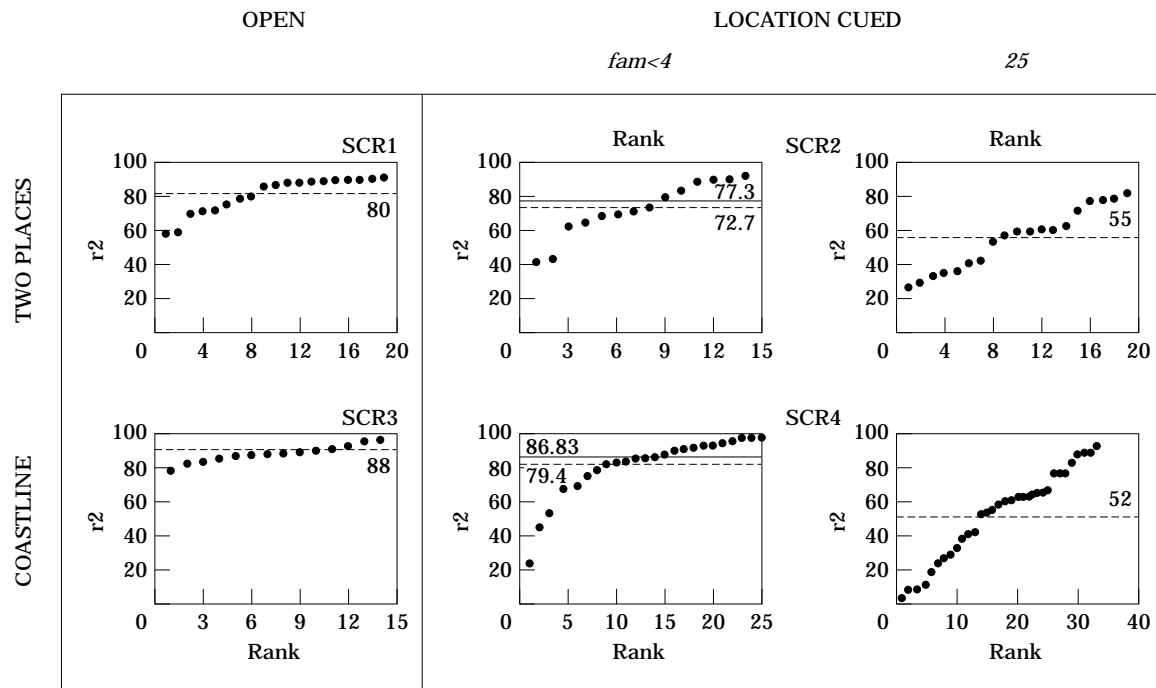
FIGURE 6.   Comparing the r^2 values from spatial cued response tests 1 to 4 (- - - =mean; — =mean-anomalies removed).

tests which use location cueing are masking the effect of spatial cueing, possibly because the location cueing leads to a number of large residuals. This masking effect can be removed using the familiarity filter. Inspection of the graphs in Figure 6 reveals that r^2 values can drop by 30–40 per cent because of location cueing, although the use of a familiarity filter will reduce this effect to only 3–20 per cent. The spatial cueing effect of a coastline raises r^2 values by 5–10 per cent.

Because the r^2 value and distortion index are so intricately linked, if the r^2 values differ between the tests so will the distortion index. It does not, however, follow that proportions of relative distortion (how the cognitive locations are relatively positioned in relation to each other) and absolute distortion (how the cognitive locations are absolutely positioned in relation to the real world locations) that make up the total distortion value will remain the same. Figure 7 shows the relationship between the absolute and relative distortion against the total amount of distortion across the spatial product. It is apparent that the spatial cueing had a significant effect upon the distortion patterns. Where there is minimal spatial cueing, absolute distortion is dominant regardless of the effects of location cueing upon the spatial product. This is reversed for the tests with detailed spatial curing where relative distortion is dominant. The detailed

spatial tests also have less total distortion. This reversal may be expected as detailed spatial cueing provides the respondent with a base scale, thus removing any absolute distortion which would be introduced by respondents misjudging the scale. Again, it might be expected that differences will occur between tests with the same spatial cueing, bur varying location cueing, due to the effect of large residuals introduced because the respondent is asked to locate unknown places.

The fact that there is a significant difference in absolute distortion between the open tests (scr1 vs scr3, $t=5·28$, $p=0·00$) and location cued tests (scr2 vs scr4, $t=5·64$, $p=0·00$), but not in relative distortion (scr1 vs scr3, $t=0·72$, $p=0·48$; scr2 vs scr4, $t=0·19$, $p=0·85$) suggests that spatial cueing only affects absolute distortion. Thus, places are still located relatively to each other to the same degree on both tests but variances in the scale of the spatial products, as a result of spatial cueing, introduces differences in absolute distortion. The effect of the spatial cueing on relative and absolute distortion becomes fully apparent when using the familiarity filter. Here both the relative ($t=2·23$, $p=0·038$) and absolute distortion ($t=4·55$, $p=0·00$) are significantly different suggesting that the relative distortion is only the same for analysis framework *25* on tests 2 and 4 because of the effects of location cueing. Again, location cueing is masking the effect of the spatial
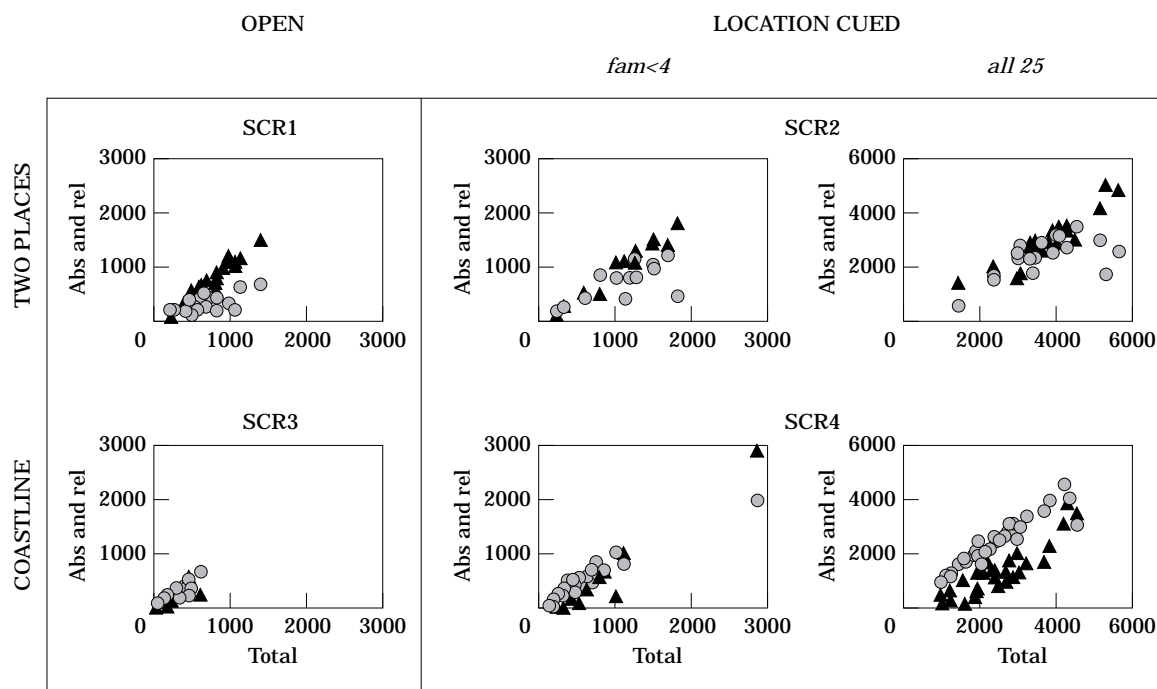
OPEN   LOCATION CUED

fam<4   *all 25*

FIGURE 7.   Comparing the absolute (▲), relative (○) and total distortion between spatial cued response tests 1 to 4.

cueing. However, unlike the r^2 data where the familiarity filter removes this masking effect, there is still a significant difference between the open and location cued tests for relative distortion. This implies that although the familiarity filter does produce results equivalent to choosing the places to locate for the r^2 results (absolutely similar), relatively they differ. This suggests that familiarity and choice are synonymous but not equivalent and that differences probably occur because of minor misreporting of familiarity.

To explore whether the location cueing affected how well respondents did in relation to each other, the r^2 values for each respondent have been ranked and plotted against each of the six analysis frameworks for spatial cued response tests 2 and 4 (Figure 8). Because framework *6* tends to cause much disturbance the results have also been displayed with this framework removed. The graphs show that, as might be expected, there are few differences between the systematic analysis frameworks *25* and *15*, but that framework *6* produces considerably different results. The reason why so much change occurs for analysis framework *6* is due to the lessening of the r^2 range and suggests that most respondents know these places to approximately the same degree. A respondent's rank is, however, much more likely to alter when using the data-defined filtering as inaccurate residuals,

caused by location cueing which reduce r^2 values, are removed. As a result, the respondent's 'true' knowledge is analysed. It appears that although some respondents are ranked consistently across the analysis frameworks, many are not. This means that how well individuals appeared to do in relation to others differs as a result of the location cueing.

In order to test the validity of the conclusions drawn from the spatial cued response tests given the low sample size, the same data were re-collected exactly 1 year later using the next set of first-year undergraduates. In this way the respondents had exactly the same amount of exposure to training and the Swansea area. Any differences between the respondents should in theory be due to weak construct validity in the tests producing poor reliability, although it is possible that differences could occur because of the abilities of the new intake, although the grades required for entry were the same as the previous year. The results for the first and second data collection were compared using two sample *t*-tests of the bidimensional regression result variables. Examination of the results revealed that there were a number of differences between the two samples. For example, the r^2 results from spatial cued response test 1 were significantly different between the two groups with the data collected a year later having dropped on average by 7·2 units. The r^2 values from the other tests,

Spatial cued response test 2   Spatial cued response test 4
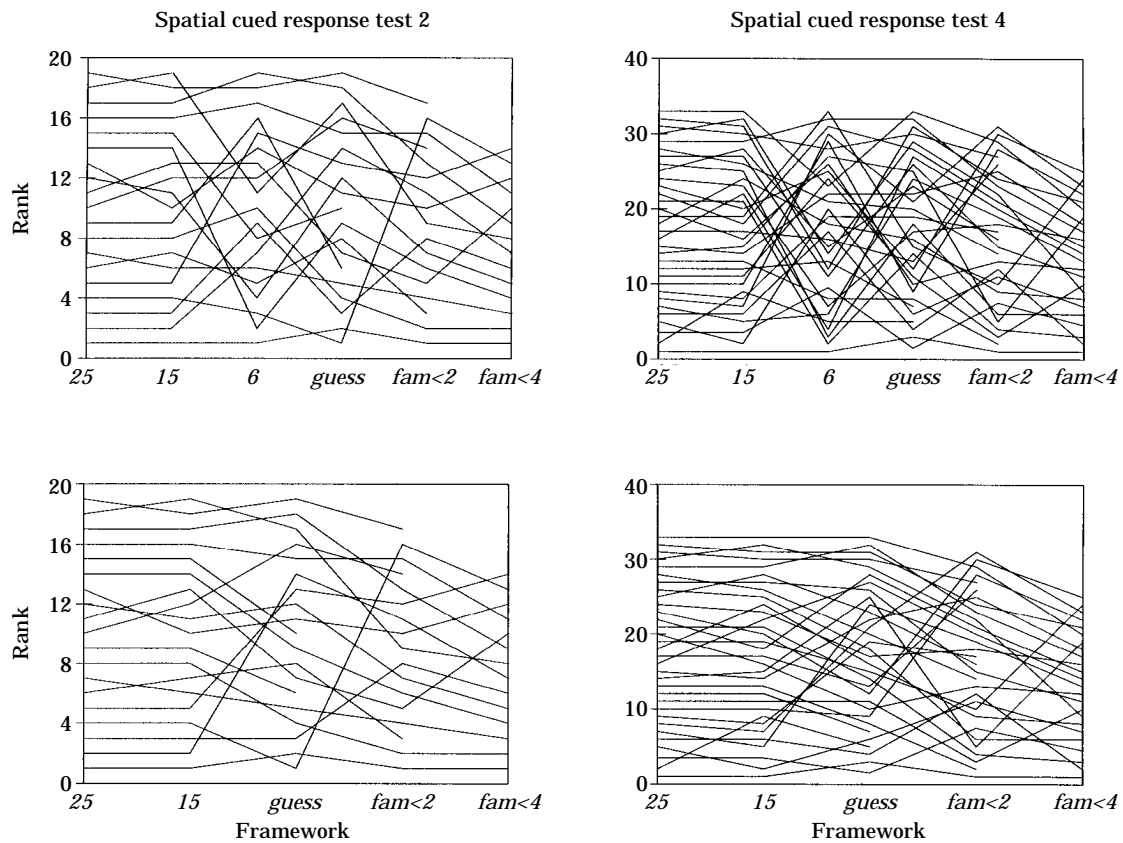

FIGURE 8.   Comparing individual rank scores across the analysis frameworks for spatial cued response tests 2 and 4.

however, did not differ significantly except for spatial cued response test 2 on analysis framework *fam <2*. Differences worth noting were the scale and angle values for spatial cued response test 2 where the second group's data needed more scaling but less rotation and there was less relative, and hence total, distortion on the first group's set of spatial cued response 3 tests.

However, the question remains concerning whether the conclusions drawn from these results would have differed. When respondent's spatial products were compared across the tests the same relationships were found to exist — respondents completing spatial cued response test 3 still produced the best results, followed by those that had completed spatial cued response test 1, those who had completed spatial cued response test 4 and lastly spatial cued response test 2. The effects of the spatial and location cueing are also still evident with the analysis frameworks improving the results. Although as noted, the tests did significantly differ on a few variables, in the main, the data corresponded well and it is fair to say that the reliability of these tests is good; they consistently

measure the same knowledge in the same way. The reason for the few differences is hard to predict. It may be the case that because the first set of respondents were used in the original pilot study, that the results from that study may have been influenced by learning. For the respondents in the second study this was the first time they had seen the tests.

Cloze procedure test 2 was analysed using a subset of the same analysis frameworks, in this case consisting of four frameworks: *15* (all the data), *guess, fam<2* and *fam<4*. In contrast to the results from the spatial cued response tests, two-tailed *t*-tests indicated that high location cueing results in less error in the spatial products than open location cueing ($t=-2.23$, $p=0.032$). It seems that cloze procedure test 2, by providing a framework of location cues with fixed spatial locations, gave the respondents the opportunity to match places to locations. Thus respondents are only required to ask of their cognitive map knowledge the question, 'is that place there?'. Cloze procedure test 1 on the other hand, although providing the exact spatial location of places, still required respondents actively to
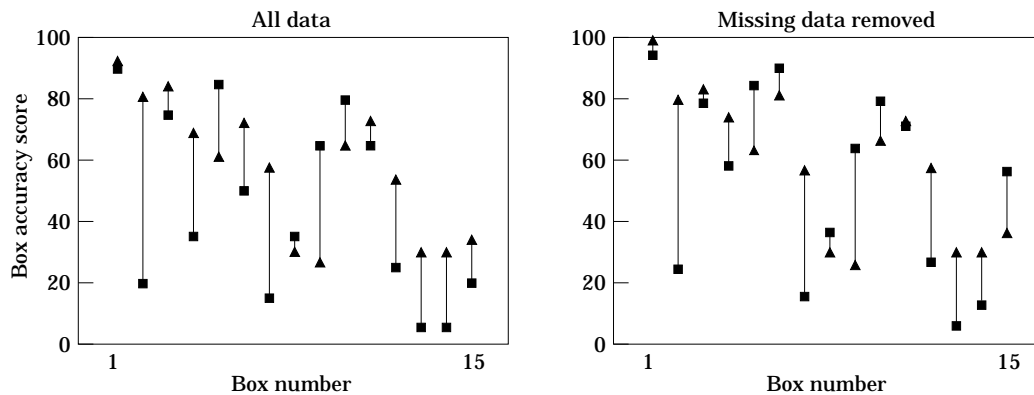
FIGURE 9.   Comparing box accuracy scores for cloze procedure tests (1 (■) and 2 (▲).

interrogate their cognitive map knowledge for the features in those locations. The question thus becomes 'which places are there?'. On spatial cued response tests 2 and 4 respondents were asked 'where is this place?' so that the location cueing provided a more demanding task because of the reduction in spatial cueing.

To draw further insight from the nature of the cloze procedure tests it is necessary to consider the results at an aggregated level by computing a measure which indicated how many times the location in each box was correctly identified. A box accuracy score is calculated by determining how many times a place has been correctly assigned to that location and dividing it by the total number of estimates. This value is then multiplied by 100 providing a value which ranges between 0 and 100. The box score was calculated with and without the missing data removed. Figure 9 shows the differences in the box accuracy scores between the two tests. It can be seen that in every box bar four the box accuracy score is higher for cloze procedure test 2. Even when the effects of the missing data have been removed the box accuracy score is better bar six boxes. It is clear that the task demands of the two tests are leading to different results. Centrally located boxes (boxes 8, 9, 10) benefit from the open-cued nature of the cloze procedure test 1, as these boxes have higher box accuracy scores than on cloze procedure test 2, especially when the effect of missing data is removed. This indicates that the two tests differ in two different ways. First, although cloze procedure test 2, on average, produces better accuracy scores, cloze procedure test 1 produces better results for centrally located, more familiar areas. The location cueing is therefore improving the box accuracy scores in unfamiliar areas by providing cues as to which locations are in those boxes. Respondents on

cloze procedure test 1, however, have the task of trying to think of a location which resides in that box. Error is being introduced to the two sets of spatial products in two different ways. On cloze procedure test 2, error is introduced by respondents recognizing that a place is located in a wrong box. On test 1 however, error is introduced by respondents recognizing that a wrong place is located in a box, although some error is due to the same method as present on cloze procedure test 2. This subtle shift in emphasis from matching boxes to places to matching places to boxes seems to be the reason for differences. Second, the missing data on cloze procedure test 2 act as a familiarity filter, so that respondents leave blank the boxes they are unfamiliar with. These were generally outlying peripheral boxes and the docks area (a place which students are unlikely to know well).

As with the spatial cued response tests and the sketch map 3 test, cloze procedure test 2 was also examined using a set of analysis frameworks. Two sample *t*-tests were calculated to compare the individual and box accuracy scores between analysis framework *15* and the other frameworks. The *t*-tests revealed that, unlike the spatial cued response tests and the sketch map 3 data, the analysis frameworks had little effect upon the accuracy scores gained. This means that the proportion of incorrect allocations to correct answers remains the same on all the analysis frameworks. The picture, however, is slightly clouded by the bias introduced into the box accuracy score by reducing the number of respondents. This gives large box accuracy scores to boxes only familiar to a few respondents. The analysis frameworks remove much unknown and guessed data, thus many of the large residuals present on analysis framework *15* have been removed by analysis framework *fam<4*. The large residuals on

the original test are introduced through a combination of miscognition and unfamiliarity. By analysis framework *fam<4*, however, it is likely that miscognition alone is responsible for the residuals. This miscognition seems to be the result of confusion caused by connected neighbours, as familiar unconnected boxes have little residual error associated with them.

Like the spatial cued response tests, data were re-collected for the cloze procedure tests exactly one year later than the first. Two sample *t*-tests of the individual accuracy scores reveal the respondents who completed the tests 1 year later performed significantly better on both tests (cz1: $t=-2.38$, $p=0.023$, $M 1=45$, $M 2=53.62$; cz2: $t=-1.81$, $p=0.080$, $M 1=57.7$, $M 2=70$). The reason for this increase is unclear, especially after the decrease on the spatial cued response tests. However, the relationship between the two tests remains the same and the box accuracy scores do not differ with the same errors occurring on both tests with the same places (cz1: $t=0.48$, $p=0.64$, $M 1=70.1$, $M 2=74.9$; cz2: $t=-1.13$, $p=0.27$, $M 1=59.4$, $M 2=70$). Although the tests did produce different individual scores, the relationships remain the same and validate the conclusions drawn from the results of the first data collection.

Investigating the results from the uni-to-multidimensional tests. It is clear from Figure 10 that the r^2 values obtained from the uni-to-multidimensional tests differed markedly from the spatial cued response tests regardless of the analysis framework used. Figure 10 shows that in all cases the spatial cued response tests produced higher r^2 values. This suggests that either the uni-to-multidimensional tests are measuring different components of configurational knowledge or are measuring the same components, but in a different way to the spatial cued response tests. The three uni-to-multidimensional tests when compared across analysis framework *6* show there are statistically significant differences between the metric MDS test and both the nonmetric MDS ($t=-8.82$, $p=0.00$) and projective convergence tests ($t=-11.45$, $p=0.00$). The nonmetric MDS and projective convergence tests seems to produce comparable r^2 results ($t=0.29$, $p=0.78$). The plot shapes, however, do differ (Figure 10). The nonmetric MDS has an *s*-shaped curve with higher numbers of top and bottom values. Alternatively, the projective convergence test produces an *s*-shaped curve where there are large numbers of middle values. This suggests that the test seems to average out the respondent's knowledge and leads to the

conclusion that respondents have very similar levels of configurational knowledge.

The analysis frameworks had no effect on the results gained from the metric MDS test. This is not surprising given the poor performance of the test. The matrix form of the test was particularly demanding requiring respondents to recognize the distances between 15 places, routes between which may not have been travelled. Even with just six places to estimate the distances between, it seems that the demands of converting configurational knowledge into 21 distance estimates are too great.

This indicates that the three uni-to-multidimensional tests differ in their ability to measure configurational knowledge, although the reasons for these differences are unclear. The spatial cued response (1 to 4) and cloze procedure tests gave the impression that they were measuring the same components of configurational knowledge but the spatial and location cueing altered the results gained. With the uni-to-multidimensional tests it is unclear whether the tests are producing different patterns of results because they are measuring different components of configurational knowledge or because the tests are measuring the same components of knowledge but with different degrees of success. If it is because of the latter, there are again two plausible reasons for difference; either the test affects cognitive processing (how the mind processes the data) or it introduces selective use of the knowledge (how the mind chooses which pieces of knowledge to use). It seems fairly intuitive that it is likely to be a combination of both of these factors. To investigate why the tests differed in the results they produced, interviews were carried out with 20 respondents. Respondents indicated that they had difficulty converting their configurational knowledge into distance and direction estimates.

Investigating the results from the recognition test. Table 5 shows how many respondents allocated each square as the correct answer on each question of the test. The top line of the table indicates that 56 respondents identified the correct location (55 knew, 1 guessed), 18 respondents thought the configuration rotated 90° was correct, the configurations rotated by 180°, 270°, inverted, inverted and rotated 270° were each selected by one respondent, and 31 (30 knew, 1 guess) thought that the configuration inverted and rotated 90° was correct. The grid highlights the residuals as those in bold and reveals facts about respondents' configurational knowledge noted in other tests. Many respondents chose squares where the configuration was rotated 90° to
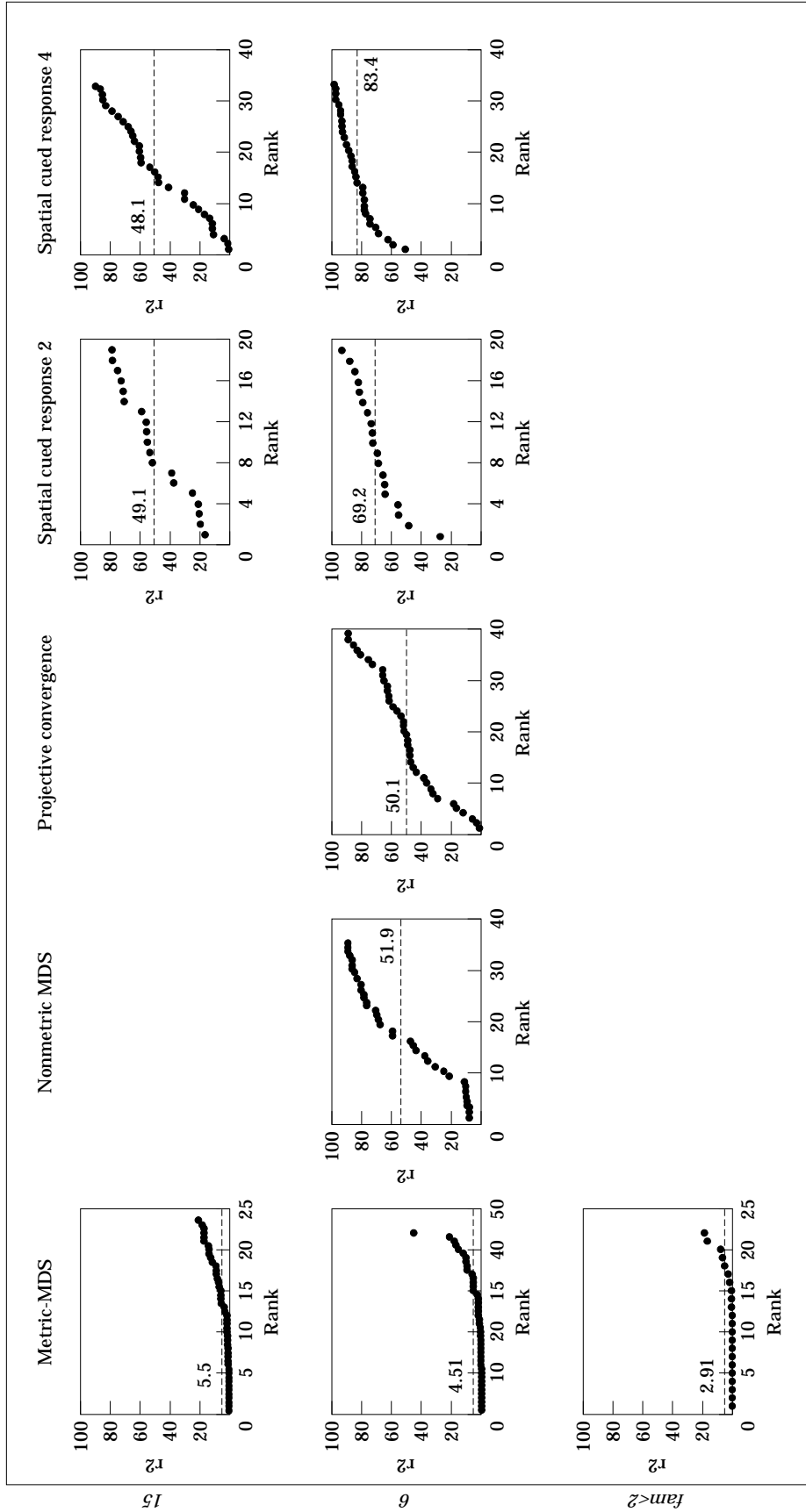
FIGURE 10.   Comparing the r² values from the uni-to-multidimensional tests and spatial cued response tests 2 and 4.

the right (column 3). This rotation was found on the low spatially cued, spatial cued response tests (1 and 2), and the objective convergence test. It is these high residuals which lead to a reduction in the configuration accuracy scores. To try and find reasons for this rotation two sample *t*-tests were calculated. Configurations containing a named place were tested against configurations not containing that place. The only significance difference was between the place indexes of configurations containing Mumbles Pier and those configurations devoid of that place. This strongly suggests that the west end of Swansea Bay (Mumbles Pier) seriously distorts the cognition of Swansea students. The bay seems to act as a topological frame of reference, whereby other places are anchored to it; if it is distorted then these other places become distorted. The east end of the bay, however, seems relatively undistorted as indicated by the high configuration accuracy score for configurations not containing Mumbles Pier. The large residuals that are not the result of the 90° rotation are the result of an alternative configuration which looked very similar to the correct configuration. This tended to happen on some configurations, especially if they had

been horizontally inverted and rotated. There were two other configurations (4 and 13) which were large residuals which did not contain Mumbles Pier. This is largely the result of two configurations looking very similar, rather then any serious miscognition.

Given the rotation effect, it may well be that in some cases the individual accuracy scores are low because of this rotation. All the individual accuracy scores were computed again, but this time the configuration rotated by 90° was taken as being correct. The results are plotted against the correct orientation in Figure 11. Sector A contains those respondents who rotated nearly every configuration 90°. Sector B contains those respondents who rotated very few of the configurations. Both sets of respondents in each sector display good consistent configurational knowledge despite sector A's knowledge being rotated. Those respondents in sector C have knowledge which is distorted in some areas, but good in others. These respondents had some configurations orientated correctly and other configurations rotated. This suggests that it is these respondents who misrecognized the western end of the Swansea Bay area in relation to the other locations.

TABLE 5
Orientation specification's aggregated results

configuration	rotation																none correct	configuration accuracy score		number of respondents	places in configs
	corr		90		180		270		inv		90		180		270				without guesses		without guesses	
	k	g	k	g	k	g	k	g	k	g	k	g	k	g	k	g	none			*		
1	55	1	**18**	0	1	0	1	0	0	1	**30**	1	0	0	1	0	0	★51·38	51·89◇	(c - 109 106)	MP SH SRS
2	79	4	**17**	0	1	0	1	0	0	0	0	0	3	0	4	0	0	76·15	75·24	(d - 109 105)	CH SLC HSV
3	68	5	**22**	2	1	0	2	0	1	0	6	1	1	0	0	0	0	66·97	67·33◇	(h - 109 101)	SLC CH MP
4	60	4	5	0	1	0	4	0	2	0	**23**	1	8	0	1	0	0	★58·72	57·69	(b - 109 104)	CH CH SRS
5	31	2	**42**	7	0	0	4	0	**16**	1	1	0	0	0	2	0	3	★30·28	32·29◇	(b - 109 96)	HSV SH MP
6	78	5	6	0	2	1	3	1	0	1	0	0	7	1	2	0	2	76·15	79·59	(f - 109 98)	HSV SLC SRS
7	17	0	**30**	2	2	0	0	0	**48**	1	1	0	1	0	4	1	2	★15·60	16·35◇	(e - 109 104)	HSV CH MP
8	33	0	**41**	5	2	0	0	0	**14**	0	**8**	0	2	0	2	0	2	★30·28	32·35◇	(b - 109 102)	SH CH MP
9	74	5	**14**	0	2	0	1	0	1	0	0	0	1	0	**9**	0	2	72·48	72·55	(a - 109 102)	HSV SH SLC
10	24	3	4	0	0	0	2	0	**8**	0	**63**	2	1	0	0	0	2	★24·77	23·53◇	(f - 109 102)	SRS SLC MP
11	62	6	**23**	0	3	0	0	0	0	0	7	3	2	0	0	0	3	62·39	63·92◇	(h - 109 97)	MP SH SLC
12	72	3	**13**	1	0	0	4	0	2	0	1	0	0	0	**9**	0	4	68·81	71·29	(e - 109 101)	SH CH HSV
13	58	1	1	0	1	0	2	0	2	1	7	0	**30**	3	2	0	1	★54·13	56·31	(e - 109 103)	CH SH SLC
14	55	4	**33**	5	0	0	3	0	1	0	1	0	1	0	1	0	5	★54·13	57·89◇	(g - 109 95)	MP HSV SLC
15	76	3	3	1	1	0	**8**	1	1	0	1	0	2	0	4	1	7	72·48	79·17	(g - 109 96)	SH HSV SRS
16	64	9	**18**	1	1	0	2	1	2	0	1	0	2	0	3	0	5	66·97	68·82◇	(c - 109 93)	MP HSV SRS
17	86	2	6	1	0	1	4	0	1	0	2	1	3	0	1	0	1	80·73	83·50	(c - 109 103)	SH SLC SRS
18	52	3	**18**	3	0	0	3	0	1	1	**17**	1	1	0	1	0	**8**	★50·46	55·32◇	(a - 109 94)	CH MP SRS
19	81	5	**9**	0	1	0	3	0	1	0	4	0	0	0	1	0	4	78·90	81·00	(h - 109 100)	SRS HSV CH
20	76	4	5	2	2	0	**8**	0	0	0	6	0	2	0	1	1	2	72·48	75·76	(f - 109 99)	SRS SLC SCH

bold = residuals > 7; ★ < 60; ◇ contained Mumbles Pier.

MP = Mumbles Pier; SH = Singleton Hospital; SRS = Swansea Railway Station; SLC = Swansea Leisure Centre;
CH = Clyne Halls of Residents; HSV = Hendrefoilan Student Village.

k = know; g = guess; none = no configuration chosen; inv = configuration horizontally inverted;
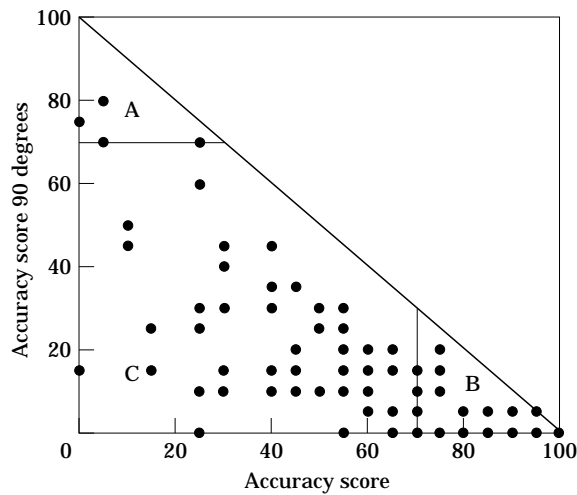* = square on test which was correct.

FIGURE 11. Comparing the accuracy scores from the correct orientation with error scores when rotated 90°.

Comparing the respondents across test groupings

Because groups of respondents completed four tests it is possible to test for convergent validity; to see whether individuals performed as well on some tests as on others. For example, if respondent A does well in comparison to his peers on test 1 does this hold true for tests 2, 3 and 4, or do some tests benefit, and others disadvantage, some respondents. In order to test the convergent validity of the tests, Spearman rank correlations were calculated for all six groups of respondents comparing individual ranks for all tests, across all analysis frameworks.

Table 6 shows the Spearman rank correlation between all the tests and analysis frameworks for group 4's tests. If all the respondents achieved the same rank on each test and framework there would be a table of perfect positive correlations. Given that some of the tests are known to introduce bias into the results it might be expected that the highest correlations will be between the tests when the familiarity filters have been used. This is not the case, with the four tests all having low correlation, negative relationships using framework *fam<4*, except for spatial cued response 4 and sketch map 3 which had a significant positive relationship ($r=0.619$). The orientation test is, however, significantly related to the other three tests when framework *25* is involved, although the metric MDS test is not related to the other two. Given the overall poor performance of the metric MDS test this is not surprising. It seems that for these particular tests when analysis framework *25* is used, respondents' rank-

TABLE 6
Spearman rank correlations for group 4 tests and analysis frameworks

	os	scr4 (*25*)	scr4 (*15*)	scr4 (*6*)	scr4 (*g*)	scr4 (*f<2*)	scr4 (*f<4*)	sk3 (*25*)	sk3 (*15*)	sk3 (*6*)	sk3 (*g*)	sk3 (*f<2*)	sk3 (*f<4*)	mds (*15*)	mds (*6*)
scr4 (*25*)	0·364														
scr4 (*15*)	0·445	0·969													
scr4 (*6*)	−0·341	0·185	0·093												
scr4 (*g*)	0·295	0·738	0·735	0·245											
scr4 (*f<2*)	−0·221	0·448	0·375	0·445	0·608										
scr4 (*f<4*)	−0·229	0·216	0·157	0·451	0·369	0·516									
sk3 (*25*)	0·659	0·699	0·710	−0·051	0·553	0·171	0·000								
ske (*15*)	0·625	0·682	0·710	−0·033	0·486	0·061	0·053	0·959							
sk3 (*6*)	0·048	0·074	0·071	0·410	0·070	0·175	−0·043	0·161	0·138						
ske (*g*)	0·381	0·469	0·488	0·037	0·505	0·257	0·107	0·781	0·694	0·444					
sk3 (*f<2*)	0·198	0·220	0·220	0·245	0·430	0·387	0·211	0·614	0·547	0·501	0·851				
sk3(*f<4*)	−0·308	0·340	0·334	0·340	0·483	0·613	0·619	0·209	0·139	0·256	0·364	0·359			
mds (*15*)	0·407	0·114	0·107	−0·015	0·085	0·072	0·258	0·100	0·120	−0·318	−0·147	−0·259	0·129		
mds (*6*)	0·543	−0·100	−0·054	−0·312	0·054	0·020	−0·045	0·318	0·215	−0·406	0·274	0·185	0·211	0·525	
mds (*f<2*)	−0·561	−0·030	−0·153	−0·032	−0·540	−0·220	−0·173	−0·316	−0·306	−0·032	−0·500	−0·669	−0·333	0·040	−0·227

Significant at 0·306.

ings tend to correlate, suggesting that those that do well on spatial cued response test 4 will have also performed well on the others. However, once the analysis frameworks are introduced, individuals radically alter their standing in relation to their peers as residual error is removed from their results. Any residual error caused by lack of familiarity cannot be removed from the orientation specification results because of the absence of a familiarity filter. The effect of the familiarity filter on sketch map 3 and spatial cued response test 4 can be seen by looking at the relationships with the orientation specification tests across the analysis frameworks. The relationships become weaker with a progression from frameworks *25* to *fam<4* (column 1).

For group 2, because the sketch map data were not quantitatively analysed but qualitatively classified the results cannot be ranked and used to compare individuals across the tests. As a result, only three tests can be compared. Like the group 1 tests the relationship between the data-defined data and the two other tests is nonsignificant (Table 7). Unlike the group 1 tests however, the relationship remains insignificant using the original analysis frameworks (*25*). This suggests that there is little relationship between these three tests, as respondents performing well on the spatial cued response test 2 could produce poor results on the other two tests. This is also true for groups 1, 3, 5 and 6 although for group 5 there is a positive significant relationship between nonmetric MDS and orientation specification and between cloze procedure 1 and orientation specification. In all, only seven out of 24 test comparisons (excluding frameworks) had positive relationships, but three of these also tested as nonsignificant in a different group. For example, the relationship for sketch map 3 and orientation specification was significant for group 4, but not group 1. The poor relationships between the tests could be caused by three factors. Either the

residuals are not being filtered from the tests with no analysis frameworks; the tests are measuring different knowledge bases; or the ability to cope with the different task demands affects the results.

Discussion

The analysis reveals that the tests produce varying results, both within and across test categories, for a number of reasons that centre upon task demands. In particular, spatial and location cueing have a dramatic impact upon the results gained. For example, on the spatial cued response tests location cueing was found to introduce large residuals into the analysis by requiring respondents to locate places they were unfamiliar with. This led to a 30–40 per cent reduction in r^2 values. Even with the introduction of analysis frameworks, designed to remove these residuals, some bias remained (3–20 per cent reduction in r^2 values). The spatial cueing provided a visual cue for respondents to 'hang' their knowledge upon and thus increased the accuracy of their spatial products (r^2 increased by 5–10 per cent). Spatial cueing was also found to have a significant effect upon the distortion patterns, with minimum spatial cueing leading to dominant absolute distortion and high spatial cueing leading to dominant relative distortion. However, it was discovered that in reality spatial cueing only affected absolute distortion which has reduced with spatial cueing (more accurate to real world locations) with relative distortion remaining the same (places still have the same relationship to each other). The relationship between spatial and location cueing was, however, found to differ between the various results. For example, location cueing was found to mask the effect of spatial cueing for the r^2, scale and angle results. This is because respondents are required to locate places they are unfamiliar with, introducing guesses (residuals) to the analysis.

TABLE 7
Spearman rank correlations for group 2 tests and analysis frameworks

	scr2 (*25*)	scr2 (*15*)	scr2 (*6*)	scr2 (*g*)	scr2 (*f<2*)	scr2 (*f<4*)	proj
scr2 (*15*)	0·981						
scr2 (*6*)	0·528	0·521					
scr2 (*g*)	0·684	0·688	0·596				
scr2 (*f<2*)	0·246	0·232	0·395	0·558			
scr2 (*f<4*)	−0·037	−0·073	0·168	0·051	0·487		
proj	0·365	0·325	0·465	0·501	0·571	0·313	
os	−0·239	−0·270	−0·280	−0·152	0·091	0·289	0·254

Significant at 0·399.

These residuals cancel out the positive effects of spatial cueing. The relationship, however, alters for the a_1 and a_2 results with spatial cueing only having an effect when location cueing is present.

The results from the cloze procedure tests, in contrast to the spatial cued response tests, indicated that the location cueing leads to less error. This was because respondents were required to match places to spaces rather than think of places to fit in spaces. The relationship between the spatial and location cueing was found to be highly complex with high location cueing masking the effect of spatial cueing, unless spatial cueing was high, whereupon location cueing became less salient. The combination of location and spatial cueing thus altered the task demands of the tests with the instructions for spatial cued response tests 1 and 3 requesting respondents to 'locate any places'; the instructions for spatial cued response tests 2 and 4 requesting respondents to answer a series of 'where is this place?' questions; the instructions for cloze procedure test 1 requesting respondents to answer a series of 'which place is here?' questions; and instructions for cloze procedure test 1 requesting respondents to answer a series of 'is that place here?' questions. These partially graphic tests were repeated a year later and, although there were a few differences between the results, the relationships between the tests remained the same, and it is fair to say that the conclusions drawn would have been very similar, thus validating the findings.

It was discovered that recognition test also suffered problems relating to spatial cueing. In an interview condition many respondents disclosed that the task of matching their knowledge to a square containing the true configuration was demanding, especially if none of the configurations matched their knowledge. It was established that most of the confusion was the result of respondents' miscognition of the location of Mumbles Pier in relation to the other locations. All the configurations containing Mumbles Pier were found to have significantly lower accuracy scores than the configurations devoid of that location. It seems that when spatial cueing is provided that does not match the respondents' knowledge the task becomes more difficult.

The influence of spatial and location cueing was highlighted by the results produced when the data was analysed using data-defined or systematic filtering. For example, the analysis frameworks for the sketch map 3 and spatial cued response tests 2 and 4 were found to have a radical effect upon the results. Not only did they affect the results but they also influenced how well individuals did in comparison to their peers. This effect upon the results was more pronounced when using data-defined filtering rather than systematic filtering. This is because the data-defined filtering specifically allowed the large residuals caused by unfamiliarity to be removed. The analysis frameworks on the cloze procedure, orientation specification and metric multidimensional scaling tests had little effect but for differing reasons. On the cloze procedure tests little use was made of the guess option, and although the familiarity frameworks did improve the accuracy scores, low numbers of respondents compensated for the effect by introducing bias. Similarly, on the orientation specification test the guess option was only used occasionally. There was little effect on the metric MDS test because the test universally produced poor results with high residual error. The error was so great that the analysis frameworks had diminutive success in removing it. The results from these frameworks suggested that familiarity is a key variable in cognitive map knowledge with scores improving when the data-defined frameworks were used.

In addition to spatial and location cueing, the task demands of some tests just proved too difficult for some respondents. For example, the uni-to-multidimensional exercises proved, in general, to be too abstract in nature, with respondents required to convert their configurational knowledge into one-dimensional distance and direction data. The r^2 values from the bidimensional regression were found to differ markedly from the spatial cued response tests. The nonmetric multidimensional scaling test led to r^2 values that were high or low. In contrast, the projective convergence test led to respondents all achieving very similar r^2 values. This suggested that the nonmetric multidimensional scaling task required certain skills and that the resection method of converting projective convergence distance and direction estimates 'averaged out' much of the variance in individual data sets. Respondents in an interview condition found that the projective convergence test did not provide enough spatial cues, with the task of conceptualizing distance and direction taxing. The metric multidimensional test was extremely abstract with respondents unable to convert their knowledge into a series of related distance estimates.

Conclusion

The study has highlighted a number of important

points. First, it has been demonstrated that task demands play a significant role in determining the results gained from a particular data collection technique. All 13 tests were shown to produce varying results both across and within test groupings. Second, no one test produces 'superior' results, with each test introducing some biases into the analysis. In the main, biases were introduced through spatial and location cueing, although some tests were too abstract and cognitively demanding. Third, that to a certain degree, biasing, such as that caused by spatial or location cueing, can be compensated for. In the present study, the effects of location cueing could be compensated against through the use of data-defined filtering. Because of these points, to increase the integrity and validity of studies concerning cognitive map knowledge it is recommended that research designs which use multiple and mutually supportive tests be adopted.

It is appreciated that the respondents used in the study were all geography students with skilled knowledge in issues of cartography. Further, the study area had a strong, misleading topography with the coastline acting as a major frame of reference. It is therefore important to replicate this work further using different populations and more mundane study areas at varying scales to gauge further the nature and influence of task demands upon the results gained from particular tests. For example, do task demands differ at various ages? Are some gender specific? How does learning affect their role? How does topography or the scale of the study area affect their role? It is important that we are aware of the limitations of the techniques we use to measure cognitive map knowledge given that we use them to test our theories and further work must be devoted to establishing both construct and convergent validity.

Acknowledgements

References

Aitken, S. C., Cutter, S. L., Foote, K. E. & Sell, J. L. (1989). Environmental perception and behavioural geography. In Wilmott, C. and Gaille, G. Eds., *Geography in America.* London: Merrill, pp. 218–238.
Anastasi, A. (1990). *Psychological Testing*, 6th Edn. New York: MacMillan.
Beck, R. J. & Wood, D. (1976*a*). Cognitive transformations from urban geographic fields to mental maps. *Environment and Behaviour,* **8,** 199–238.
Beck, R. J. & Wood, D. (1976*b*). Comparative developmental analysis of individual and aggregated cognitive maps of London. In G. T. Moore & R. G. Golledge, Eds., *Environmental Knowing.* Stroudsberg, PA: Dowden, Hutchinson and Ross, pp. 173–184.
Boyle, M. J. & Robinson, M. E. (1978). Cognitive mapping and understanding. In D. T. Herbert & R. J. Johnston, Eds., *Geography and the Urban Environment: Progress in Research and Applications,* **2.** London: Wiley, pp. 59–82.
Bryant, K. J. (1984). Methodological convergence as an issue within environmental cognition research. *Journal of Environmental Psychology,* **4,** 43–60.
Burroughs, W. & Sadalla, E. (1979). Asymmetries in distance cognition. *Geographical Analysis,* **11,** 414–421.
Buttenfield, B. P. (1986). Comparing distortion on sketch maps and MDS configurations. *Professional Geographer,* **38,** 238–246.
Cadwallader, M. T. (1979). Problems in cognitive distance and their implications to cognitive mapping. *Environment and Behaviour,* **11,** 559–576.
Cronbach, L. J. (1990). *Essentials of Psychological Testing.* New York: Harper Collins.
Evans, G. W., Fellows, J., Zorn, M. & Doty, K. (1980). Cognitive mapping and architecture. *Journal of Applied Psychology,* **65,** 474–478.
Evans, G. W. & Pezdek, K. (1980). Cognitive mapping: knowledge of real-world distance and location information. *Journal of Experimental Psychology: Human Learning and Memory,* **6,** 13–24.
Gerber, R. & Kwan, T. (1994). A phenomenological approach to the study of pre-adolescents' use of maps in a wayfinding exercise in a surburban environment. *Journal of Environmental Psychology,* **14,** 265–280.
Golledge, R. G. (1977). Environmental cues, cognitive mapping and spatial behaviour. In Burke, D. *et al.,* Eds., *Behaviour — Environment Research Methods.* Institute for Environmental studies, University of Wisconsin, Wisconsin, pp. 35–46.
Golledge, R. G. (1992). Place recognition and wayfinding: making sense of space. *Geoforum,* **23,** 199–214.
Golledge, R. G., Briggs, R. & Demko, D. (1969). The configuration of distances in intraurban space. *Proceedings of Association of American Geographers,* **1,** 60–65.
Golledge, R. G., Gale, N. & Richardson, G. D. (1987). Cognitive maps of cities II: studies of selected populations. *National Geographical Journal of India,* **33,** 1–16.
Hardwick, D. A., McIntyre, C. W. & Pick, H. L. (1976). The content and manipulation of cognitive maps in children and adults. *Monographs of the Society for Research in Child Development,* **41,** 1–55.
Hart, R. (1979). *Children's Experience of Place.* New York: Irvington.
Howard, R. B., Chase, S. D. & Rothman, M. (1973). An analysis of four measures of cognitive maps. In Preisner, W. F. E. Ed., *Environmental Design*

Research, **1.** Stroudsberg, PA: Dowden, Hutchinson and Ross.

Humphreys, J. S. (1990). Place learning and spatial cognition: a longitudinal study of urban newcomers. *Tijdschrift Voor Econmicshe en Sociale Geografie,* **81,** 364–380.

Jacobson, R. D. (1992). Spatial cognition through tactile mapping. *Swansea Geographer,* **29,** 79–88.

Kirasic, K. C., Allen, G. & Siegel, A. W. (1984). Expression of configurational knowledge of large-scale environments: student's performance of cognitive tasks. *Environment and Behaviour,* **16,** 687–712.

Kirasic, K. C., Siegel, A. W. & Allen, G. (1981). The development of basic processes in cognitive mapping: recognition in context memory. *Child Development,* **51,** 302–305.

Kitchin, R. M. (1994*a*). Cognitive maps: what are they and why study them? *Journal of Environmental Psychology,* **14,** 1–19.

Kitchin, R. M. (1994*b*). Spatial familiarity as a variable in cognitive mapping. *Swansea Geographer,* **31,** 42–52.

Kitchin, R. M. (1995). Issues of validity and integrity in cognitive mapping research: investigating cognitive mapping. Unpublished PhD: University of Wales Swansea.

Lloyd, R. (1989). Cognitive maps: encoding and decoding information. *Annals of the Association of American Geographers,* **79,** 101–124.

Lockman, J. J., Rieser, J. J. & Pick, H. L. (1981). Assessing blind travellers' knowledge of spatial layout. *Journal of Visual Impairment and Blindness,* **75,** 321–326.

Lynch, K. (1960). *The Image of the City.* MA: MIT Press.

MacKay, D. B. (1976). The effect of spatial stimuli on the estimation of cognitive maps. *Geographical Analysis,* **8,** 439–451.

Magana, J. R., Evans, G. W. & Romney, A. K. (1981). Scaling techniques in the analysis of environmental cognition data. *Professional Geographer,* **33,** 294–301.

Matthews, M. H. (1984). Cognitive maps: a comparison of graphic and iconic techniques. *Area,* **16,** 33–40.

Montello, D. R. (1991). The measurement of cognitive distance: methods and construct validity. *Journal of Environmental Psychology,* **11,** 101–122.

Murphy, J. (1978). Measures of map accuracy assessment and some early Ulster maps. *Irish Geography,* **11,** 89–101.

Passini, R. & Proulx, G. (1988). Wayfinding without vision: an experiment with congenitally blind people. *Environment and Behavior,* **20,** 227–252.

Pearce, P. L. (1981). A study of travellers perceptions of a section of countryside. *Journal of Environmental Psychology,* **1,** 141–155.

Pocock, D. C. D. (1976). Some characteristics of mental maps: an empirical study. *Transactions of the Institute of British Geographers,* **1,** 493–512.

Robinson, M. E. (1974). Cloze procedure and spatial comprehension test. *Area,* **9,** 137–142.

Saarinen, T. F., MacCabe, C. L. & Morehouse, B. (1988). Sketch maps of the world as surrogates for world geographic knowledge. Discussion paper 83-3 Dept of Geography and Regional Development, Univ. of Arizona, Tucson.

Spencer, C. & Darviezeh, Z. (1981). The case for developing a cognitive environmental psychology that does not underestimate the abilities of young children. *Journal of Environmental Psychology,* **1,** 21–31.

Stea, D. & Blaut, M. (1973). Some preliminary observations on spatial learning in school children. In R. M. Downs & D. Stea, Eds., *Image and Environment.* Chicago: Aldine, pp. 226–234.

Thorndyke, P. W. & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology,* **14,** 560–589.

Tobler, W. R. (1965). Computation of the correspondence of geographic patterns. *Papers and Proceedings of the Regional Science Association,* **15,** 131–139.

Tobler, W. R. (1976). The geometry of mental maps. In R. G. Golledge & G. Rushton, Eds., Columbus, OH: Ohio State University Press, pp. 69–82.

Ungar, S., Blades, M. & Spencer, C. (1995). Visually impaired children's strategies for memorizing a map. *British Journal of Visual Impairment,* **13,** 27–32.

Vanetti, E. J. & Allen, G. L. (1988). Communicating environmental knowledge — the impact of verbal and spatial abilities on the production and comprehension of route directions. *Environment and Behavior,* **26,** 667–682.

Waterman, S. & Gordon, D. (1984). A quantitative-comparative approach to analysis of distortion in mental maps. *Professional Geographer,* **36,** 326–337.

Wood, D. & Beck, R. (1976). Talking with Environmental A: An experimental mapping language. In G. T. Moore & R. G. Golledge, Eds., *Environmental Knowing.* Stroudsberg, PA: Dowden, Hutchinson and Ross, pp. 351–361.