European Journal of Orthodontics 30 (2008) 503–507
doi:10.1093/ejo/cjn007
Advance Access publication 16 July 2008

Is the Royal London Space Analysis reliable and does it influence orthodontic treatment decisions?

M. Al-Abdallah, J. Sandler and K. O'Brien
School of Dentistry, University of Manchester, UK

SUMMARY The purpose of this study was to investigate the reliability of the Royal London Space Analysis (RLSA) and to evaluate its influence on orthodontic treatment decisions. Thirty-one case records were collected to represent various levels of crowding and different types of malocclusions. Seventeen examiners assessed these records and completed a data sheet that recorded information on their treatment decision. One month later, the examiners attended a course on the RLSA and then used the analysis to rescore the 31 cases. The models were also scored by the expert who led the course and these were then considered the 'gold standard' scores. After a further month, the examiners reapplied the RLSA and formulated a treatment plan for each set of patient records. A paired Student's *t*-test and intraclass correlation coefficient (ICC) were used to assess the agreement in scoring RLSA, a paired sample *t*-test was used to compare the scores with the gold standard, and finally the reliability in treatment planning was determined using kappa (κ) statistics.

The scores for lower arch crowding showed the highest inter-examiner agreement with an ICC of 0.93 whereas the lowest level of agreement was for upper arch space requirement with an ICC of 0.77. Intra-examiner agreement was generally high, particularly for the assessment of lower arch crowding (ICC = 0.93) and lower arch space requirement (ICC = 0.88). There was excellent validity for all the examiners against the gold standard scores with a paired samples correlation ranging between 0.96 for lower arch crowding and 0.79 for upper arch space requirement. The intra-examiner reliability in treatment decision was only moderate, with an average κ value of 0.52 (maximum 0.82, minimum 0.24). Intra- and inter-examiner agreement for scoring the RLSA was acceptable. Nevertheless, the additional information obtained from the application of the RLSA did not have a substantial impact on the treatment decisions.

Introduction

Orthodontic treatment planning should follow a detailed and accurate assessment of space requirement and anchorage needs. Ideally, there should be a uniform, universally accepted method of measuring these components. Space assessment varies between a visual evaluation of patient records to the detailed application of measurements that apply scores to the various components of the malocclusion which either provide space or require space for their correction. This latter approach has the aim of justifying extraction and anchorage decisions and reducing variation in possible treatment plans.

Over recent years, several space analyses have been developed and advocated as tools that may be used in treatment planning (Herren *et al.*, 1973; Rudge, 1982; Bhatia and Harrison, 1987; Harris *et al.*, 1987; Richmond, 1987; Schirmer and Wiltshire, 1997). One of these is the Royal London Space Analysis (RLSA) (Kirschen *et al.*, 2000a,b). This was one of the first analyses to include measurements for most of the components of a malocclusion including: crowding, spacing, space gained or required from arch width changes, antero-posterior change in incisor position, and angulation and inclination of the maxillary incisors. The first section of the analysis is followed by summing these measurements to produce a total space requirement for the upper and the lower arches. These measurements are then integrated with the space requirements of possible treatment mechanics such as tooth reduction or enlargement, extraction, space opening for prosthetic replacement, and mesiodistal molar movement (Figure 1). The benefits of this approach are to provide consistency in treatment planning decisions and act as an aid to trainee orthodontists so that they fully appreciate space requirements and their influence on treatment decisions.

While this analysis is gaining popularity, there have been no published studies that have evaluated its reliability and validity particularly with respect to any influence on orthodontic treatment decisions. Thus, this was the aim of the present study.

The null hypotheses tested were

1. The RLSA does not have acceptable validity or reliability.
2. There is no effect of RLSA on orthodontic treatment decisions.

Materials and methods

Sample

Using data from previous investigations (Han *et al.*, 1991; Keeling *et al.*, 1996; Ribarevski *et al.*, 1996; Pair *et al.*, 2001), it was calculated that 31 sets of patient records were required

ROYAL LONDON HOSPITAL - ORTHODONTIC SPACE PLANNING

Space requirements:

+ = Space available or gained

- = Space required or lost

	LOWER		UPPER
Crowding and spacing:		mm	
		mm	
Leveling occlusal curve:		mm	
		mm	
Arch width change:		mm	
		mm	
Incisor A/P change:		mm	
		mm	
Angulation/inclination change:		mm	
		mm	
	TOTAL	mm	
		mm	
Space creation/utilization in addition to any planned above			
Tooth reduction/enlargement: (+ or -)		mm	
		mm	
Extractions:	+	mm	+
	mm		
Space opening for prosthetic replacement:	-	mm	-
	mm		
Molar distal movement:	+	mm	+
	mm		
Molar mesial movement:	-	mm	-
	mm		
Differential U/L growth: (+ or -)		mm	
		mm	
RESIDUE (should = 0)		mm	
		mm	

Figure 1 Royal London Space Proforma. Reprinted from the American Journal of Orthodontics and Dentofacial Orthopedics (Kirschen *et al*., 2000a), Copyright 2000, with permission from Elsevier.

to be evaluated by 17 orthodontic examiners to conduct a study with an alpha level of 0.05 and beta value of 0.1. This would give sufficient power to detect a 3-mm change in the scores of RLSA and a 15 per cent change in the overall treatment decisions for each examiner. The 17 examiners included 10 orthodontic specialist and seven postgraduate students in the first and second years of their speciality programme.

Materials

Thirty-one sets of records comprising pre-treatment study models, coloured facial and intra-oral photographs, dental pantomograms, lateral cephalograms, and their tracings were made available. The cases were selected so that the sample was representative of different types of malocclusion (Table 1) and different levels of crowding severity according to Proffit (2000).

Procedure

This study was carried out in three main stages as shown in Figure 2. The 17 participants were asked to examine the 31 sets of records and complete a data sheet that recorded information on their treatment decision, extraction pattern, and the anchorage method used (Figure 3). One month later, the examiners attended a course on the RLSA given by a clinician who routinely used the analysis (JS). The examiners then applied the space analysis to the 31 sets of study casts and recorded their findings. Finally, after a further one-month period, they reapplied the RLSA to the same study casts and repeated their treatment decision for each set of patient records.

Statistical analysis

The data were checked for normality and simple summary statistics were produced using the Statistical Package for Social Sciences version 11.5 (SPSS Inc., Chicago, Illinois, USA) and Stata version 8 (Stata Corporation, College Station, Texas, USA) statistical software programs.

The agreement of each examiner's scores with the gold standard scores of the expert was calculated using a paired sample *t*-test. Inter- and intra-rater reliability in scoring RLSA was evaluated with the intraclass correlation coefficient (ICC) and the paired Student's *t*-test. The intra-reliability of the treatment decision was evaluated with kappa (κ) statistics. An overall κ value across all examiners was also computed.

Finally, so that any effect of the application of the analysis on the treatment decisions of the examiners could be measured, it was decided that the examiner's decision was most likely to be influenced by the upper and lower arch space requirements. These requirements were divided into high (when the space required was 8 mm or more) and low (when it was less than 8 mm). The treatment decisions were then treated as a dependent variable and the effect of space requirements and

Table 1 The characteristics of the records in the study.

Crowding (%)				Incisor classification (%)				
None	Mild ≤4 mm	Moderate ≥5 and ≤9 mm	Severe ≥10 mm	Class I	Class II division 1	Class II division 2	Class II intermediate	Class III
12.9	16.1	45.2	25.8	22.6	38.7	16.1	12.9	9.7

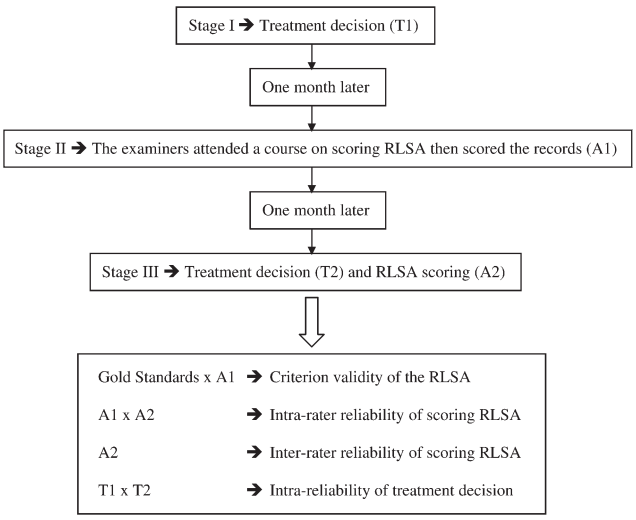

Figure 2 Flow chart illustrating the procedure and analysis used.

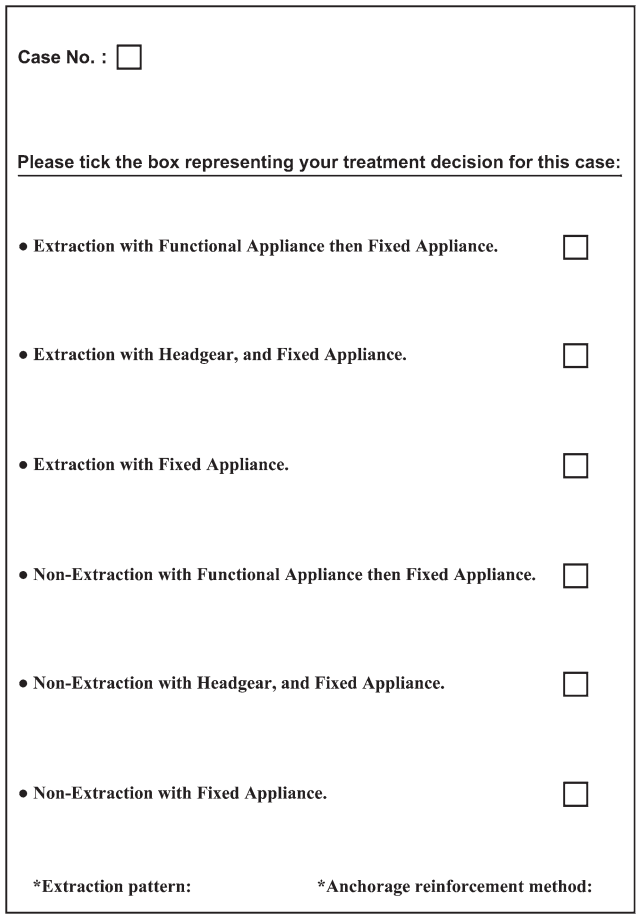

Figure 3 The data sheet used in the current investigation to record the examiners' treatment decisions.

possession of an orthodontic qualification were evaluated using a binary logistic regression. The analysis was run for each examiner.

Results

The agreement of the examiners with the gold standard is shown in Table 2. The correlation ranged between 0.96 (maximum 0.98, minimum 0.93) for lower arch crowding and 0.79 (maximum 0.90, minimum 0.57) for total upper space requirement.

The inter- and intra-examiner agreement when using the space analysis, represented by the ICC, is shown in Table 3. This reveals that inter-examiner agreement ranged from 0.93 for the assessment of lower arch crowding to 0.77 for total upper space requirement. When intra-examiner agreement was considered, this ranged from 0.93 for lower arch crowding to 0.68 for total upper space requirement.

The data on intra-examiner reliability of the examiners' treatment decisions before and after application of the space analysis are shown in Table 4. The average κ value for all examiners was 0.52 (0.24–0.82) representing moderate agreement.

The impact of the RLSA on the examiners' treatment decisions was measured using regression analysis. This showed that the RLSA did not systematically influence their decisions in the use of headgear, anchorage reinforcement, extractions, or functional appliances. It did, however, indicate that some individual examiners were influenced for some decisions but this was not consistent, neither was it explained by whether the individual possessed an orthodontic qualification.

Table 2 Agreement of the examiners with the 'gold standard' using paired samples *t*-test correlations.

Measurements	Paired samples correlations	Maximum and minimum
Lower arch crowding	0.96	0.98–0.93
Total lower arch space requirement	0.90	0.97–0.81
Upper arch crowding	0.88	0.95–0.46
Total upper arch space requirement	0.79	0.90–0.57

Table 3 Inter- and intra-rater coefficients of agreement for scoring Royal London Space Analysis among the 17 examiners as represented by the intraclass correlation coefficient (ICC).

Measurements	ICC (inter-rater agreement)	ICC (average intra-rater agreement)	Maximum and minimum
Lower arch crowding	0.93	0.93	0.98–0.83
Total lower arch space requirement	0.87	0.88	0.95–0.77
Upper arch crowding	0.87	0.85	0.97–0.42
Total upper arch space requirement	0.77	0.68	0.95–0.21

Table 4 Intra-rater reliability of treatment decisions made by the 17 examiners as represented by kappa.

Examiner	Kappa	Standard error	95% confidence interval
1	0.51	0.12	0.28–0.75
2	0.63	0.12	0.41–0.86
3	0.50	0.14	0.23–0.77
4	0.37	0.11	0.15–0.59
5	0.54	0.11	0.37–0.76
6	0.67	0.11	0.46–0.88
7	0.24	0.13	0.01–0.40
8	0.79	0.10	0.59–0.78
9	0.72	0.12	0.48–0.96
10	0.50	0.12	0.27–0.73
11	0.82	0.08	0.66–0.99
12	0.55	0.11	0.33–0.78
13	0.38	0.12	0.15–0.61
14	0.36	0.12	0.14–0.59
15	0.32	0.12	0.08–0.55
16	0.30	0.11	0.07–0.52
17	0.58	0.10	0.38–0.78

Discussion

The results of this study reveal that the RLSA had consistent reliability both between and within examiners. It was found that its use did not influence the relevant treatment decision of the orthodontists when applied as part of the treatment planning process. Nevertheless, the level of calibration against the gold standard was high and suggests that it is possible to easily apply and teach the RLSA.

Scoring and measurements

It is well established that there is an inherent error in the application of all orthodontic clinical measurements, and the error varies depending on the feature being measured. For example, it is not surprising that the assessment of crowding, which is carried out by direct measurement, is accurate (Schirmer and Wiltshire, 1997). However, other measurements such as estimating the curve of Spee, assessment of space gained by arch expansion, space required for incisors anterior/posterior position change, and angulation/inclination change were less accurate as they rely on the examiner's judgement which is based on their own experience and knowledge.

Inter-examiner reliability in scoring the RLSA

When these data were evaluated in depth, it was found that scoring lower arch crowding had the highest inter-examiner agreement whereas the lowest agreement was for upper arch space requirement. These results are in agreement with those reported by Pair *et al.* (2001) where 30 orthodontists were asked to assess different aspects of malocclusion from the study casts of 10 patients. Among these variables, they assessed the reliability of scoring maxillary and mandibular arch crowding, which was 89 and 93 per cent, respectively.

Intra-examiner reliability in scoring the RLSA

In most orthodontic studies, intra-examiner agreement has been shown to be generally high (Keeling *et al.*, 1996; Luke *et al.*, 1998; Pair *et al.*, 2001). This was the case in the present investigation, particularly for the assessment of lower arch crowding and lower space requirement. It also appears that reliability was lower for the total compared with crowding as a single variable, and this was due to accumulation of the error in scoring these variables when summing the total.

Intra-examiner reliability in treatment decisions

The main aim of the application of a space analysis is to provide information that may be used as a guide to treatment decisions. It was postulated that the application of the RLSA would lead to a change in the examiners' decisions, following the addition of new information, and that this would be consistent. This was, however, not the case since their decisions taken without the analysis did not substantially change following its application.

Previous investigations have shown that orthodontic treatment decisions are somewhat unreliable and the results of the present study confirm this finding. An example is the research by Han *et al.* (1991) which looked at the consistency of orthodontic treatment decisions relative to diagnostic records. In that study, five orthodontists examined the records of 15 Class II division 1 patients twice with a 4 to 6 week interval. The intra-examiner reliability ranged from 53 to 73 per cent, with an average of 65 per cent.

In a similar study, Ribarevski *et al.* (1996) asked 10 orthodontists to evaluate the records of 60 Class II division 1 patients twice with a 1 month interval. The average κ score for intra-examiner agreement was 0.69 (maximum 0.96, minimum 0.54). The two studies may not be strictly comparable because the investigation was confined to subjects with Class II division 1 malocclusions which may have resulted in less variation in treatment decisions.

The impact of RLSA on treatment decisions

The results of this study revealed that the use of the RLSA did not influence most of the orthodontists' decisions. It appears that the assessment of space requirements alone does not influence the treatment decision in a simple linear relationship. This may be explained by considering that decision making and treatment planning are complex procedures involving several factors which are both external and personal to the clinician. This has been outlined by Kay and Nuttall (1995a,b) who stated that 'sources of variation in clinical decisions include the clinician's perception of the probability of success of treatment, patient–dentist relationship, patient attendance, risk–benefit ratio, dentist's and patient's values of dental health care, dentist's personal treatment threshold, and finally patient's financial ability'.

The process of treatment planning is also influenced by innate variability of the clinician that arises from two main

sources. These are perceptual and judgemental variations. Perceptual variation is when people 'see' things differently. This is influenced by past experiences and factors such as poor lighting, whereas, judgemental variation is when a clinician decides on different treatment plans based on their personal opinion, experience, and the evidence base that underpin treatment options.

The results in this current investigation showed good reliability in scoring all components of the RLSA. Thus, the panel in this study agreed on what they saw and there was minimal perceptual variation. As a result, judgemental variation was likely to be the main source of the disagreement in their treatment decision.

In the present study, the patient was assumed to be 'the ideal patient' who is highly co-operative, with excellent oral hygiene and no relevant medical history. This assumption eliminated the judgemental variations produced from patient factors and left those related to the examiner and the surrounding environment as the main source of disagreement between the examiners. These factors may possibly include examiner knowledge and his/her basic training followed by personal experience.

The process of treatment planning is far from straight forward taking into consideration that orthodontic treatment is characterized by an almost complete lack of high-level evidence for the effectiveness of various treatment methods. In this respect, it is not surprising that there was such a marked variation between the orthodontists in the current study despite good agreement in the application of the RLSA. There is no doubt that attempts should be made to reduce this variation mainly by increasing the amount of research that is carried out into the effects of orthodontic treatment.

Conclusions

The following conclusions may be drawn from this study:

1. The RLSA had good reliability.
2. There was no impact of the use of the RLSA on the treatment decision-making process.
3. The use of RLSA might have reduced the perceptual variability among orthodontists but had no effect on the judgemental variation, which was the main source of disagreement in the process of orthodontic treatment decision making.

Address for correspondence

M. Al-Abdallah
Orthodontic Department
School of Dentistry
University of Manchester
Higher Cambridge Street
Manchester M15 6FH
UK
E-mail: mariam.al-abdallah@postgrad.manchester.ac.uk

Acknowledgement

The authors thank the orthodontic consultants who participated in the study: D. Bearn, S. Chadwick, S. Caldwell, N. Mandall, M. Read, D. Lewis, P. Banks, and C. Low. In addition, we are grateful to the postgraduate orthodontic students who took part in the study. Finally, we would like to thank Professor H. Worthington for the statistical advice.

References

Bhatia S N, Harrison V E 1987 Operational performance of the travelling microscope in the measurement of dental casts. British Journal of Orthodontics 14: 147–153

Han U K, Vig K W, Weintraub J A, Vig P S, Kowalski C J 1991 Consistency of orthodontic treatment decisions relative to diagnostic records. American Journal of Orthodontics and Dentofacial Orthopedics 100: 212–219

Harris E F, Vaden J L, Williams R A 1987 Lower incisor space analysis: a contrast of methods. American Journal of Orthodontics and Dentofacial Orthopedics 92: 375–380

Herren P, Schmoker R, Jordi T 1973 Arch shape and space balance determined by arcogramme technique. Transactions of the European Orthodontic Society pp. 61–73

Kay E, Nuttall N 1995a Clinical decision making—an art or a science? Part II: making sense of treatment decisions. British Dental Journal 178: 113–116

Kay E, Nuttall N 1995b Clinical decision making—an art or a science? Part V: patient preferences and their influence on decision making. British Dental Journal 178: 229–233

Keeling S D, McGorray S, Wheeler T T, King G J 1996 Imprecision in orthodontic diagnosis: reliability of clinical measures of malocclusion. Angle Orthodontist 66: 381–391

Kirschen R H, O'Higgins E A, Lee R T 2000a The Royal London Space Planning: an integration of space analysis and treatment planning: part I: assessing the space required to meet treatment objectives. American Journal of Orthodontics and Dentofacial Orthopedics 118: 448–455

Kirschen R H, O'Higgins E A, Lee R T 2000b The Royal London Space Planning: an integration of space analysis and treatment planning: part II: the effect of other treatment procedures on space. American Journal of Orthodontics and Dentofacial Orthopedics 118: 456–461

Luke L S, Atchison K A, White S C 1998 Consistency of patient classification in orthodontic diagnosis and treatment planning. Angle Orthodontist 68: 513–520

Pair J W, Luke L, White S, Atchinson K, Englehart R, Brennan R 2001 Variability of study cast assessment among orthodontists. American Journal of Orthodontics and Dentofacial Orthopedics 120: 629–638

Proffit W R 2000 Contemporary orthodontics. Mosby, St Louis

Ribarevski R, Vig P, Vig K D, Weyant R, O'Brien K 1996 Consistency of orthodontic extraction decisions. European Journal of Orthodontics 18: 77–80

Richmond S 1987 Recording the dental cast in three dimensions. American Journal of Orthodontics and Dentofacial Orthopedics 92: 199–206

Rudge S J 1982 A computer program for the analysis of study models. European Journal of Orthodontics 4: 269–273

Schirmer U R, Wiltshire W A 1997 Manual and computer-aided space analysis: a comparative study. American Journal of Orthodontics and Dentofacial Orthopedics 112: 676–680