Original Article

# Intra-examiner and Inter-examiner Reproducibility in Irregularity Index Measurements

Ersat Cem İrezli[1] (ID), Mücahid Faik Şahin[2] (ID), Rana Demir[3] (ID), Aslı Baysal[2] (ID)

[1]Private Practice, İzmir, Turkey
[2]Department of Orthodontics, İzmir Katip Çelebi University School of Dentistry, Ankara, Turkey
[3]Private Practice, Konya, Turkey

## ABSTRACT

**Objective:** This study aimed to assess intra-examiner and inter-examiner reproducibility in irregularity index measurements.

**Methods:** Twenty plaster casts of moderately crowded arches were randomly selected, and five contact point displacements (CPD) at lower anterior segment (through no. 33 to no. 43) were recorded using digital calipers on two different time points by three examiners to determine irregularity index (sum of five CPDs). To evaluate intra-examiner and inter-examiner differences, paired t-test and analysis of variance were used, respectively. Correlation analyses were performed between examiner pairs, and intra-class correlation coefficients (ICC) were determined. Statistical significance was set at p≤0.05.

**Results:** Only a few of the repeated measurements of examiners showed significant differences (p≤0.05). All researchers were consistent in repeated measurements (p=0.000), and ICCs ranged between 0.916 and 0.986. For one CPD measurement, a statistically significant difference was detected among examiners (p=0.020). High correlation was found for inter-examiner repeatability (p<0.05), and ICCs ranged between 0.739 and 0.984. But when the difference of 1.5 mm among measurements was set as clinically relevant, the percentages of these values for repeated measures were 15%, 5%, and 45% for examiner 1, 2, and 3, respectively. These percentages were 25%, 80%, and 65% for examiner pairs.

**Conclusion:** Irregularity index may be a misleading index to determine anterior alignment especially when measuring small CPD.

**Keywords:** Irregularity index, reproducibility, tooth crowding, dental model

## INTRODUCTION

The pretreatment incisor crowding is considered an important factor for post-retention stabilization (1-3). In the daily clinical practice, it is subjectively recorded but rarely measured. A method called as Little's Irregularity Index (LII) was developed by Little (4) to measure the misalignment of the mandibular anterior teeth. The contact point displacements (CPDs) of six mandibular anterior teeth were measured, and the total amount of displacements provided a total score of irregularity (4). This method is simple and reliable to calculate the "quantitative score of mandibular alignment," but it is insufficient to define the severe single tooth displacement (4, 5). LII was accepted as a valid outcome measure to evaluate retention procedures according to Cochrane Collaboration's review (6).

Little's Irregularity Index has long been used to evaluate the post-retention stability; and recently, the orthodontists have extended its inclusion for maxillary arch to evaluate the effects of various brackets (7, 8), retainers (9-11), and treatment modalities (11, 12).

The reproducibility of the LII has been tested by comparing two separate occasion of measurements recorded by different examiners (4, 8, 13). The results of the studies (8, 13, 14) showed significantly high correlations. However, in some studies (15, 16), the index was reported not to yield reliable results, and consequently no consensus was

**Table 1.** Statistical comparison of intra-examiner and inter-examiner differences

| | Examiner 1 | | | | | | Examiner 2 | | | | | | Examiner 3 | | | | | | ANOVA |
| | T1 | | T2 | | Difference | | T1 | | T2 | | Difference | | T1 | | T2 | | Difference | | |
| CPDs | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33–32 | 1.81 | 1.15 | 2.05 | 1.22 | −0.24*** | 0.27 | 1.89 | 1.14 | 2.05 | 1.15 | −0.16* | 0.34 | 2.55 | 1.20 | 2.73 | 1.45 | −0.18 | 0.58 | ns |
| 32–31 | 1.63 | 1.01 | 1.79 | 1.03 | −0.16** | 0.24 | 1.78 | 0.98 | 1.87 | 0.99 | −0.09 | 0.38 | 2.08 | 1.25 | 2.25 | 1.17 | −0.17 | 0.47 | ns |
| 31–41 | 1.02 | 0.85 | 1.09 | 0.83 | −0.07 | 0.27 | 0.98 | 0.85 | 1.03 | 1.01 | −0.04 | 0.42 | 1.36 | 1.05 | 1.52 | 1.22 | −0.15 | 0.36 | ns |
| 41–42 | 1.41 | 0.92 | 1.46 | 0.92 | −0.05 | 0.26 | 1.29 | 0.93 | 1.43 | 0.98 | −0.13 | 0.34 | 1.66 | 1.07 | 1.9 | 1.13 | −0.24* | 0.42 | ns |
| 42–43 | 1.71 | 0.72 | 1.91 | 0.69 | −0.17* | 0.35 | 1.91 | 0.71 | 2.14 | 0.73 | −0.23** | 0.36 | 2.14 | 0.83 | 2.7 | 1.02 | −0.55*** | 0.51 | 0.011 (1 vs. 3, 2 vs. 3) |
| Sum | 7.62 | 2.96 | 8.31 | 2.88 | −0.69*** | 0.71 | 7.87 | 2.69 | 8.54 | 2.87 | −0.66** | 0.82 | 9.81 | 3.34 | 11.11 | 3.39 | −1.30*** | 1.30 | ns |

*p < 0.05; **p < 0.01; ***p < 0.001; ns: non-significant
CPDs: Contact point displacements

reached in the literature. Therefore, this study aimed to evaluate the repeatability of the LII measurements performed on two separate occasions by three independent examiners.

## METHODS

Ethical approval was obtained from the research ethics committee of the İzmir Katip Çelebi University (no. 210). A sample size of 20 models was calculated based on 90% statistical power, 0.10 effects size, 0.05 type I error, and 0.85 intra-class correlation coefficient for three raters.

The sample was selected from initial records of patients at Orthodontic Clinic of İzmir Katip Çelebi University. The plaster models were selected according to the following inclusion criteria: (1) fully erupted teeth, (2) moderately crowded mandibular arch, (3) no previous orthodontic treatment, and (4) no missing mandibular teeth (excluding third molars). The casts of patients with craniofacial syndromes or developmental dental anomalies were excluded.

The impression protocol in our clinic was as follows: the mandibular study casts were recorded using alginate impressions (Cavex Tulip, Haarlem, The Netherlands). Alginate was manually mixed according to the manufacturer's instructions. Impressions were rinsed with cold water and then disinfected for 10 min. Dental hard plaster (Denstone; Heraeus Kulzer, South Bend, IN) was poured into the impression and left to harden for 45 min. The casts were removed from the impression and stored at room temperature (22°C±1°C). Samples were given a number to ensure that the patient would not be identified and measurements would be made with a blinded manner.

Three examiners-senior orthodontic residents (E.C.I, R.D, and M.F.S)-who were working independently recorded the measurements on the casts. Each plaster model was measured twice with 1-month interval by the same examiner. The CPDs of the six anterior lower teeth were measured with a digital Vernier caliper.

Special care was taken to measure only the horizontal linear displacements between each CPD by holding the caliper parallel to the occlusal plane based on the protocol defined by Little (4). Vertical displacements between CPDs were not evaluated.

### Statistical Analysis
All statistical analyses were performed using the Statistical Package for Social Sciences version 22.0 (IBM Corp, Armonk, NY, USA). Statistical significance was set at p<0.05. Data was normally distributed.

The repeatability of measurements for the same examiner was evaluated using ICC, and the mean differences between two measurements were calculated with paired *t*-test. Reproducibility was evaluated with intra-class correlation coefficient (ICC) that is defined as "the proportion of the total variance due to the between subject variance" (17). The comparison among three examiners for CPD differences was performed using analyses of variance (ANOVA).

161

| **Table 2.** Intra-class correlation coefficients for intra-examiner reproducibility and inter-examiner reliability | | | | | | |
|---|---|---|---|---|---|---|
| | **Contact points** | | | | | **Sum** |
| | 33-32 | 32-31 | 31-41 | 41-42 | 42-43 | |
| **Intra-examiner reproducibility** | | | | | | |
| First examiner | 0.986 | 0.986 | 0.973 | 0.979 | 0.928 | 0.985 |
| Second examiner | 0.978 | 0.96 | 0.947 | 0.967 | 0.933 | 0.978 |
| Third examiner | 0.950 | 0.961 | 0.974 | 0.962 | 0.916 | 0.961 |
| | | | | | | |
| **Inter-examiner reliability** | | | | | | |
| First measurements | 0.969 | 0.972 | 0.970 | 0.961 | 0.875 | 0.975 |
| Second measurements | 0.963 | 0.971 | 0.971 | 0.941 | 0.875 | 0.971 |
| All ICC's were statistically significant at p=0.000 level | | | | | | |
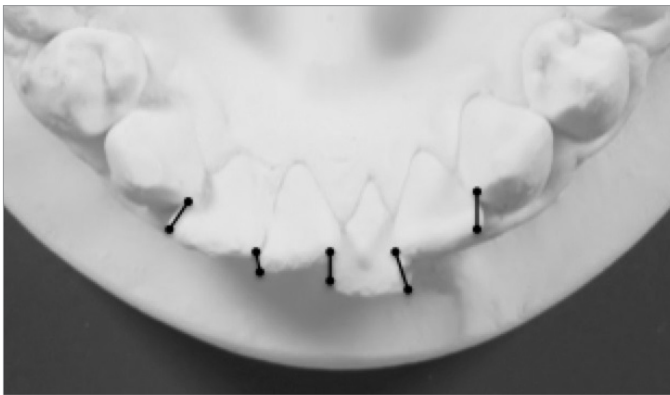


**Figure 1.** The measurement of contact point displacements

## RESULTS

The mean, standard deviation, and mean differences for each of the five CPD (LR 3/2, LR 2/1, L 1/1, LL 1/2, and LL 2/3) measurements for two different time points are given in Table 1. Three out of five measurements for the first examiner and two out of five measurements for the second and third examiners were found to be different between first and second measurements. For all examiners, the sum of CPDs (LII) between time points was found to be different and statistically significant. The differences were 0.69 mm, 0.66 mm, and 1.30 mm for the first, second, and third examiner, respectively. The one-way ANOVA showed statistically significant difference for one measurement (LR 3/2) among examiners. The differences between first and second, and second and third examiners were statistically significant (p=0.011).

Intra-class correlation coefficients for intra-examiner reliability ranged between 0.916 and 0.986 (Table 2). Thus, high intra-examiner reproducibility was recorded (ICC>0.750). Also, ICC for inter-examiner reliability ranged between 0.739 and 0.984. ICCs for sum of CPDs (irregularity index) were higher than 0.750.

The difference more than 1.5 mm between repeated measurements was assumed as clinically relevant; 15% measurements of examiner 1, 5% measurements of examiner 2, and 45% measurements of examiner 3 were more than 1.5 mm. These were 25%, 80%, and 65% for inter-examiner measurements.

## DISCUSSION

Several authors have reported high correlation coefficients for repeatability of LII measured on casts by different examiners (8, 14, 16, 18, 19). Similar to the above-mentioned studies, high ICCs were found for the repeated measures of the same investigator and the measurement of investigator couples in this study.

To generalize an acceptable level of orthodontic reliability, Roberts and Richmond (20) suggested that an ICC value of R below 0.4 is poor reliability, between 0.4 and 0.75 is fair to good, and above 0.75 is excellent. In this study, intra-examiner and inter-examiner measurements were found to be statistically correlated (ICC>0.75), while only one value indicates good and the others are excellent for inter-examiner reliability (Table 2). On the other hand, even if the correlations were found high, when the CPDs more than 1.5 mm was assumed as clinically relevant, the percentages of examiner reproducibility values were found relatively variable (Table 2). The correlation coefficients showed a linear relationship between two measurements, but it does not necessarily indicate a direct relationship (13). Therefore, the high correlation coefficients reported in the literature may be deceptive in interpreting examiner reliability. Macauley et al. (16) evaluated coefficients of variation, and reported significant differences among examiners.

In this study, five CPDs that constitute the LII score were statistically analyzed separately different from the literature (4, 8, 13, 14, 18, 21). As the LII has a cumulative nature, single CPD measurement errors magnified when the total amount was calculated as LII score. Sjögren et al. (13) evaluated the reliability of LII between two examiners who had been given written manuals and 8 h of calibration of the measuring technique before study. They found differences and variability between examiners, and reported that the technique is not an appropriate tool to measure the irregularity (13). As the experience of using the index increases, further studies may be performed to find if the rate of error will decrease.

The Little's Irregularity Index evaluation on conventional plaster casts may be challenging because of the use of hand-held cali-

pers that need to be kept parallel to the occlusal plane. In addition, the researcher brought the anatomical tooth contact points "by eye" in line with the caliper tips, without using magnification. If the contact point cannot be reached due to crowding, determination of the contact point as the most probable point may lead to individual differences in measurements. With the advances in digital technology and orthodontics, plaster casts started to be replaced by digital study models (22, 23). Treatment results have been predictably determined with great accuracy using intra-oral scanners (24).

Almasoud and Bearn (14) compared the LII reliability between photographic and cast model assessments, and they showed excellent inter-examiner reliability. They concluded that measurements made on photographic images were reliable and repeatable. They encountered problems with study models like storing models, random assignment, measuring overlapped contact points, and wearing of the model surfaces (14).

The correlation of examiners with ICC was found to be excellent in this study. However, when the difference of 1.5 mm among measurements was set as clinically relevant, the examiners showed different ratios between their own measurements and among themselves. Therefore, according to the repetitive measurements, irregularity index may be a misleading index to determine anterior alignment especially when measuring small CPDs. Photographic and 3D model assessment with LII may be more useful to determine anterior displacement. Future researches are needed to evaluate the accuracy of LII on photographs and digital models.

LII should be performed with great caution to determine anterior alignment especially when measuring small (<0.05) CPDs. The examiner training could be an important factor, and it must be taken into account.

## CONCLUSION

Within the limitations of this study, following conclusions may be drawn;

* Correlation of reproducibility of the LII was found statistically high between inter-examiners and intra-examiners.
* The reproducibility of the LII of small CPDs is low.
* Irregularity index may be a misleading index to determine anterior alignment especially when measuring small CPD.

**Author Contributions:** Concept - M.F.Ş., A.B.; Design - M.F.Ş., A.B., C.İ., R.D.; Data Collection and/or Processing - M.F.Ş., A.B., C.İ., R.D.; Analysis and/or Interpretation - A.B., C.İ.; Literature Search - A.B., C.İ., M.F.Ş.; Writing Manuscript - C.İ., A.B.; Critical Review - M.F.Ş., A.B., C.İ., R.D.

## REFERENCES

1. Edwards JG. A long-term prospective evaluation of the circumferential supracrestal fiberotomy in alleviating orthodontic relapse. Am J Orthod Dentofac Orthop 1988; 93: 380-7. **[CrossRef]**
2. Reitan K. Tissue rearrangement during retention of orthodontically rotated teeth. Angle Orthod 1959; 29: 105-13.
3. Surbeck BT, Årtun J, Hawkins NR, Leroux B. Associations between initial, posttreatment, and postretention alignment of maxillary anterior teeth. Am J Orthod Dentofac Orthop 1998; 113: 186-95. **[CrossRef]**
4. Little RM. The irregularity index: a quantitative score of mandibular anterior alignment. Am J Orthod 1975; 68: 554-63. **[CrossRef]**
5. Little RM. Stability and relapse: early treatment of arch length deficiency. Am J Orthod Dentofac Orthop 2002; 121: 578-81. **[CrossRef]**
6. Littlewood SJ, Millett DT, Doubleday B, Bearn DR, Worthington HV. Retention procedures for stabilising tooth position after treatment with orthodontic braces. Cochrane Database Syst Rev 2006; 1: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858. CD002283.pub4/full. **[CrossRef]**
7. Pandis N, Polychronopoulou A, Eliades T. Active or passive self-ligating brackets? A randomized controlled trial of comparative efficiency in resolving maxillary anterior crowding in adolescents. Am J Orthod Dentofac Orthop 2010; 137: 12-3. **[CrossRef]**
8. Wahab RM, Idris H, Yacob H, Ariffin SH. Comparison of self-and conventional-ligating brackets in the alignment stage. Eur J Orthod 2012; 34: 176-81. **[CrossRef]**
9. Rowland H, Hichens L, Williams A, Hills D, Killingback N, Ewings P, et al. The effectiveness of Hawley and vacuum-formed retainers: a single-center randomized controlled trial. Am J Orthod Dentofac Orthop 2007; 132: 730-7. **[CrossRef]**
10. Thickett E, Power S. A randomized clinical trial of thermoplastic retainer wear. Eur J Orthod 2010; 32: 1-5. **[CrossRef]**
11. Destang D, Kerr W. Maxillary retention: is longer better? Eur J Orthod 2003; 25: 65-9. **[CrossRef]**
12. Sadowsky C, Schneider BJ, BeGole EA, Tahir E. Long-term stability after orthodontic treatment: nonextraction with prolonged retention. Am J Orthod Dentofac Orthop 1994; 106: 243-9. **[CrossRef]**
13. Sjögren AP, Lindgren JE, Huggare JÅ. Orthodontic study cast analysis-reproducibility of recordings and agreement between conventional and 3D virtual measurements. J Digit Imaging 2010; 23: 482-92. **[CrossRef]**
14. Almasoud N, Bearn D. Little's irregularity index: photographic assessment vs study model assessment. Am J Orthod Dentofac Orthop 2010; 138: 787-94. **[CrossRef]**
15. Harris EF, Vaden JL, Williams RA. Lower incisor space analysis: a contrast of methods. Am J Orthod Dentofac Orthop 1987; 92: 375-80. **[CrossRef]**
16. Macauley D, Garvey TM, Dowling AH, Fleming GJ. Using Little's Irregularity Index in orthodontics: Outdated and inaccurate? J Dent 2012; 40: 1127-33. **[CrossRef]**
17. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. Stat Med 2001; 20: 3205-14. **[CrossRef]**
18. Bernabé E, Flores-Mir C. Estimating arch length discrepancy through Little's Irregularity Index for epidemiological use. Eur J Orthod 2006; 28: 269-73. **[CrossRef]**

19. Dowling AH, Burns A, Macauley D, Garvey TM, Fleming GJ. Can the intra-examiner variability of Little's Irregularity Index be improved using 3D digital models of study casts? J Dent 2013; 41: 1271-80. [CrossRef]

20. Roberts C, Richmond S. The design and analysis of reliability studies for the use of epidemiological and audit indices in orthodontics. Br J Orthod 1997; 24: 139-47. [CrossRef]

21. Little RM, Riedel RA, Artun J. An evaluation of changes in mandibular anterior alignment from 10 to 20 years postretention. Am J Orthod Dentofac Orthop 1988; 93: 423-28. [CrossRef]

22. Syrek A, Reich G, Ranftl D, Klein C, Cerny B, Brodesser J. Clinical evaluation of all-ceramic crowns fabricated from intraoral digital impressions based on the principle of active wavefront sampling. J Dent 2010; 38: 553-9. [CrossRef]

23. Stevens DR, Flores-Mir C, Nebbe B, Raboud DW, Heo G, Major PW. Validity, reliability, and reproducibility of plaster vs digital study models: comparison of peer assessment rating and Bolton analysis and their constituent measurements. Am J Orthod Dentofac Orthop 2006; 129: 794-803. [CrossRef]

24. Quimby ML, Vig KW, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based digital models. Angle Orthod 2004; 74: 298-303.