

A Critique of the Index of Complexity, Outcome and Need



Name: **Dominique Ferreira**

Student Number:

Date of Submission: **14 June 2005**

**A CRITIQUE OF THE INDEX OF COMPLEXITY, OUTCOME
AND NEED**

BY

DOMINIQUE ABERGAIL FERREIRA

A mini-thesis submitted in partial fulfilment of the requirements for the degree of Magister Scientiae Dentium in Orthodontics in the Faculty of Dentistry, University of the Western Cape.

Date of Submission: June 2005

SUPERVISOR:  **Dr. M. Ferguson:** Consultant, Department of Orthodontics, Faculty of Dentistry and WHO Oral Health Collaborating Centre, University of the Western Cape.

CO-SUPERVISOR: **Prof. N. Myburgh:** Department of Community Oral Health, Faculty of Dentistry and WHO Oral Health Collaborating Centre, University of the Western Cape.

KEY WORDS

INDICES

MALOCCLUSION

ICON

GOLD STANDARD

TREATMENT NEED

COMPLEXITY

TREATMENT OUTCOME

DEGREE OF IMPROVEMENT

VALIDITY

RELIABILITY



LIST OF ABBREVIATIONS

AAO	American Association of Orthodontics
COCSTOC	Commission on Classification and Statistics for Oral Conditions
DAI	Dental Aesthetic Index
DFI	Dento-facial Index
DHC	Dental Health component
FDI	Fédération Dentaire Internationale
HLD	The Handicapping Labio-lingual Deviation Index
HMAR	Handicapping Malocclusion Assessment Record
ICON	Index of Complexity. Outcome and Need
IOTN	Index of Orthodontic Need
MI	Malalignment Index
MSE	Malocclusion Severity Estimate
OFI	Occlusal Feature Index
OI	Occlusal Index .
PAR	Peer Assessment Rating
SCAN	Standardized Continuum of Aesthetic Need
TPI	Treatment Priority Index .
WHO	World Health Organization

ABSTRACT

The development of a uniform method of epidemiological assessment and grading of malocclusion has been of interest for several decades. Recently, Daniels and Richmond (2000) proposed a new orthodontic index namely the Index of Complexity, Outcome and Need (ICON). Their aim was to develop a single index for assessing treatment inputs and outcomes.

The aim of this study was to critique the ICON and to assess to the extent to which each component of the ICON fulfils the ideal requirements of the ideal index as identified in a World Health Organization Report (WHO, 1966). The study was performed in three parts: 1) a gold standard was established to test reliability and validity of the ICON; 2) to assess ease of use and simplicity of the index; 3) and to test the applicability of the index on patients and study casts.

The results showed that the ICON identified 25% of the cases as 'no treatment', as apposed to the 100% of the gold standard. Validity of the index was shown to be 'poor' for complexity ($\kappa = 0.2$) and degree of improvement ($\kappa = 0.34$) and 'excellent' for outcome. Reliability was high for all the components except for treatment need ($\kappa = 0.63$).

This study concluded that except for *complexity* and *degree of improvement*, the index performed well with respects to reliability, validity (of treatment outcome), ease of use and simplicity and applicability to patient and study casts.

DECLARATION

I, Dominique Abergail Ferreira, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

SIGNED

.....

D.A.Ferreira

Date Day of June of 2005



ACKNOWLEDGEMENTS

1. To all the participants of this study, the orthodontists, the dentists, the oral hygienist and the patients who agreed to be part of this research.
2. To Dr Maurice Ferguson, for his kindness, mentorship and lateral thinking. His support will always be remembered. And to his family who sacrificed their family time without which the completion of this project would not have been possible.
3. To Prof Neil Myburgh, for his assistance, encouragement and statistical advice.
4. To my husband, for his dual parenting, unending patience, support and continued brilliance.
5. To my precious daughter, Rebecca, for adapting so well to the absence of her mother.
6. To my Mother, for her love and sacrifices.
7. To my parents in-law, for their love and support.

DEDICATION

To my soul mate and love of my life, my husband for his understanding and for allowing me to pursue my passion; my beautiful daughter for being such a sweet little girl in her mother's absence.



TABLE OF CONTENTS

KEY WORDS.....	I
LIST OF ABBREVIATIONS.....	II
ABSTRACT.....	III
DECLARATION.....	IV
ACKNOWLEDGEMENTS.....	V
DEDICATION.....	VI
LIST OF TABLES.....	X
CHAPTER ONE: INTRODUCTION.....	1
CHAPTER TWO: LITERATURE REVIEW.....	4
INTRODUCTION.....	5
REQUIREMENTS OF AN IDEAL INDEX.....	5
TYPES OF INDICES.....	9
DIAGNOSTIC CLASSIFICATIONS.....	10
Angles Classification.....	10
Ackerman and Proffit System.....	11
EPIDEMIOLOGICAL INDICES.....	13
Index of Tooth Position.....	13
The Dentofacial Index.....	14
The Malalignment Index.....	15
The Occlusal Feature Index.....	15
Björks' Method.....	16
Method For Measuring Occlusal Traits.....	17
A Quantitative Assessment Of Occlusal Features.....	18

INDICES OF TREATMENT NEED.....	18
The Handicapping Labio-Lingual Deviation Index	18
Maryland and California Modifications	20
Malocclusion Severity Estimate	21
Treatment Priority Index.....	22
Handicapping Malocclusion Assessment Record	23
Occlusal Index	25
The Swedish System	26
The Dental Aesthetic Index.....	28
The Standardized Continuum of Aesthetic Need	30
Index of Orthodontic Need	31
TREATMENT OUTCOME INDICES	34
The Peer Assessment Rating Index.....	35
INDICES OF TREATMENT COMPLEXITY	39
COMPARATIVE STUDIES	39
THE INDEX OF COMPLEXITY, OUTCOME AND NEED	42
CHAPTER THREE: AIM, OBJECTIVES AND METHODOLOGY.....	46
AIM	47
OBJECTIVES	47
METHODOLOGY.....	47
Study Design.....	47
The Sample	47
Data analysis.....	48
Materials	48
Method	48

Statistical Analysis	51
Ethical Statement	52
CHAPTER FOUR: RESULTS.....	53
CHAPTER FIVE: DISCUSSION.....	70
CHAPTER SIX: CONCLUSION	80
REFERENCES	82
APPENDIX	96
APPENDIX I.....	97
APPENDIX II	105
APPENDIX III.....	106



LIST OF TABLES

Table 1. Levels of severity of a malocclusion as established by the Malocclusion Severity Estimate.	22
Table 2. Components and weights of the Dental Aesthetic Index.....	29
Table 3. Logistic regression weightings for treatment need and outcome	44
Table 4. The suggested subdivisions of Cohen’s Kappa statistic.....	52
Table 5. Distribution of IOTN (DHC) grades in sample (n = 100).....	54
Table 6. Inter-rater agreement of the Experts with the Gold Standard.....	55
Table 7. Intra-examiner Reliability (n = 25)	56
Table 8. Calibrated vs Specialist and Non-specialist	65
Table 9. A comparison of Specialist vs Non-Specialist ICON scores.....	66
Table 10. Patient ICON Score vs Cast ICON score	69
Table 11. ICON scoring protocol	97
Table 12. ICON index variables, weightings and cut-off values for treatment need and outcome decisions	102
Table 13. ICON index complexity cut-off values	104
Table 14. Degree of improvement Scale.....	104
Table 15. A 2-by-2- contingency Table of Decisions to treat or not to treat for ICON vs. Expert Opinion	106
Table 16. Calculation of validity.....	107
Table 17. Example of data collection sheet	108

LIST OF FIGURES

Figure 1. Ackerman and Proffit System.....	12
Figure 2. Child in position ready for orthometric measurement (Elsasser, 1953).....	14
Figure 3. The Index of Treatment Standards nomogram	36
Figure 4. Treatment Need	57
Figure 5. Complexity	58
Figure 6. Treatment Outcome.....	59
Figure 7. Degree of Improvement.....	60
Figure 8. Complexity for 'No treatment' cases.	62
Figure 9. Treatment Outcome 'No treatment' cases	63
Figure 10. Degree of Improvement 'No treatment' cases.....	64
Figure 11. Treatment need. Patient vs cast.....	67
Figure 12. Complexity. Patient vs Cast.....	68

CHAPTER ONE: INTRODUCTION



The traditional orthodontic diagnosis is a qualitative, descriptive procedure unsuited to a quantitative evaluation of treatment need. As a result, several systems of assessing malocclusion and evaluating treatment need have been developed in the past 50 years. These indices are procedures that generate and summarize data about the malocclusion and return a numeric value (Firestone et al., 2002).

The development of a uniform method of epidemiological assessment and grading of malocclusion has been of interest for several decades. Due to the multiplicity of measurement methods and the difficulty in standardising criteria, Baume (1970) expressed concerns about the lack of suitable methods of recording malocclusion. Jago (1974), in a review of 45 studies of malocclusion in 18 countries, reported similar difficulties when comparing his findings. The inability to develop a universal occlusal index can be traced to an increasing understanding of the multifactorial nature of malocclusion.

Indices, like fashion, go through trends. In the sixties the Handicapping Labio-Lingual Deviation Index was popular, in the seventies it was the Occlusal Index, the eighties saw the rise of the Index of Complexity and Treatment Need and Dental Aesthetic index and in the nineties the Peer Assessment Rating index became the index of choice.

Recently, Daniels and Richmond (2000) proposed a new orthodontic index namely the Index of Complexity, Outcome and Need (ICON). Their aim was to develop a single index for assessing treatment inputs and outcomes. The authors claim the index to be valid for both treatment need, complexity and outcome assessments in as much as it represents a broadly based international body of expert orthodontic opinion. This index is intended for use in late mixed dentition and permanent dentition and may be applied clinically to patients and to casts without modification. They describe the index as simple to use taking approximately one minute to apply, requiring only a millimeter ruler and the Aesthetic Component Scale of the IOTN.

Few studies of validity have been done on the ICON (Firestone, 2002; Savastano, 2003). An assessment of the extent to which this newly developed index to the ideal fulfills the requirements of an index, as described by the WHO (1966), is also lacking.

This study therefore aims to do a critical analysis of the ICON.

CHAPTER TWO: LITERATURE REVIEW



INTRODUCTION

An occlusal index is a numerical scale that is derived by scoring specific characteristics of a malocclusion to assess how far it deviates from an ideal occlusion (Richmond, 2001). Occlusal indices offer objectivity in the recording of traits of malocclusion in numerical or categorical formats, thereby reducing the reliance on subjective assessments (Buchanan et al., 1993). The use of indices in orthodontics allows a more uniform application and interpretation of criteria for treatment need and treatment induced changes (McGuinness and Stephens, 1995).

Since Angle classified malocclusion in 1899, there have been many attempts to develop an occlusal index that can be used to record the prevalence of malocclusion,  treatment need, priority, outcome and complexity. There is still no occlusal index that can assess all these aspects (Turner, 1990).

REQUIREMENTS OF AN IDEAL INDEX

In a World Health Organization Report on international methodology for epidemiological studies of oral diseases the following requirements for an ideal index had been identified (WHO, 1966) namely:

1. Validity
2. Reliability
3. Validity over time
4. Speed of application

5. Simplicity
6. Clinical relevance
7. Adaptability
8. Applicability
9. Acceptability

1. Validity is the ability of an index to measure that which it purports to measure (Carlos, 1970; Summers, 1971; McGuinness and Stephens, 1995; Roberts and Richmond; 1997; Beglin et al., 2001). In a clinical or epidemiological context the assessment of validity takes place against a gold standard; which is derived from the expert opinions of a group of orthodontists (Beglin et al., 2001). Subjective assessment by experienced clinicians offers one basis for the analysis of the performance of an index (Summers, 1971; Turner, 1990; Roberts and Richmond, 1997; Hamden and Rock, 1999; Daniels and Richmond, 2000; Beglin et al., 2001; Firestone, 2002; Savastano, 2003).

According to Kaey (1993) true validation based on a gold standard is impossible. However, there are no other standards available in assessing malocclusions.

2. Reliability, also referred to as reproducibility or precision, is the extent to which a measurement is repeatable under identical conditions. The term intra-examiner reliability refers to the consistency of repeated observation

by an observer, whilst inter-examiner reliability relates to observations being consistent amongst a group of observers (Carlos, 1970; Summers, 1971; Shaw et al. 1991; Buchanan et al., 1993; McGuinness and Stephens, 1995; Roberts and Richmond; 1997; Beglin et al., 2001).

3. The index should be valid during time. The index should consider the normal development of occlusion (Summers, 1971). According to Tarvit and Freer (1998) there is a need to distinguish between developmental features and persistent traits of malocclusion.

Summers (1971) proposed that the developmental changes in occlusal disorders may consist of either a basic *orthodontic defect* or a *symptom of developmental change*. A *basic orthodontic defect* is defined as a constant occlusal dysfunction which may exist before, during and after the development of occlusion. This defect may be:

- a) skeletal, such as the size of the mandible being disproportionate to the maxilla,
- b) dental, such as a discrepancy in the size of the teeth and the jaw,
- c) neuromuscular, such as a tongue-thrust or
- d) combinations of the above.

A symptom of a developmental change is defined as an adaptation to development, for example the flaring and spacing of the maxillary permanent incisors normally seen in the early mixed dentition. The index

must concentrate on, and be sensitive to, the basic defect and must not be unduly sensitive to the symptom.

For an index to be valid over a period of time, the index score of the occlusal disorder should either remain constant or increase over that period (Summers, 1971; Turner, 1990).

4. The examination required should be performed quickly, even by examiners without special instruction in orthodontic diagnosis (Shaw et al., 1991). The examination procedure should require a minimum of judgement and requisite equipment and the instruments should be practical in the actual field situation (WHO, 1966).



5. It should also be simple (McGuinness and Stephens, 1995), accurate and yield itself to modification for the collection of data (Draker, 1960). The Index value should be amendable to statistical analysis (Jamison and McMillan, 1966; Draker, 1960; Tang and Wei, 1993).

6. The score should correspond closely with the clinical importance of the disease stage it represents. The status of the group is expressed by a single number which corresponds to a relative position on a finite scale with definite upper and lower limits; running by progressive gradation from zero (absence of disease), to the ultimate point (disease in its terminal stage). The index should be equally sensitive throughout the scale.

7. It should be facile enough to permit the study of a large population without undue cost in time or energy.

8. The index should be applicable both clinically and to study casts (Draker, 1960).

9. The index should be acceptable to the profession and public alike (McGuinness and Stephens, 1995; Abdullah and Rock, 2001).

It has proved difficult to devise a single index that fulfils all these criteria and this has led to the proliferation of different methods (Abdullah and Rock, 2001).



TYPES OF INDICES

A review of various types of Indices will be considered under five main headings as described by Otuyemi and Jones (1995), Shaw et al. (1995) and Abdullah and Rock (2001) namely:

1. Diagnostic Classification
2. Epidemiological
3. Treatment need
4. Treatment success
5. Treatment complexity.

According to Shaw (1995) it is the purpose rather than the content of an index that categorises it.

DIAGNOSTIC CLASSIFICATIONS

Diagnostic classifications are descriptive and enable malocclusions to be categorised. They tend to be qualitative rather than quantitative, which makes them of limited value as research tools.

Angles Classification

The most widely used classification is that proposed by Edward H. Angle (1898). He proposed that if the mesiobuccal cusp of the maxillary first molar articulates in the buccal groove of the mandibular first molar, and if the rest of the teeth in the arch are aligned an ideal occlusion will result. Angle described three basic types of what he termed malocclusion, all of which represented deviations in an anteroposterior plane.



There have been many critiques of Angle's classification of malocclusion. One of the most severe critics was Calvin Case (1921) who pointed out that Angle's method disregarded (in treatment planning as well as classification) the relationship of the teeth to the face; and although malocclusion was a three-dimensional problem with the Angle system only anteroposterior deviations were taken into consideration.

Another criticism of the Angle classification was that it merely described the relationship of the teeth and did not differentiate between dentoalveolar and skeletal discrepancies. Ackerman and Proffit (1969) also found that the classification does not indicate the complexity of the problem.

When the reliability of Angles classification was tested by Gravely and Johnson (1974), they found that the inter- and intra-examiner error levels in categorizing Angle Class II, Division 2 malocclusions, were both relatively high. They further suggested that the classification was unreliable because of the difficulty associated with asymmetry between left and right sides, or where tooth movements had occurred because of factors such as crowding and premature loss of deciduous teeth.

Other criticisms include its inability to attach a value to, and to express other characteristics of malocclusion (Otuyemi and Jones, 1995).

Despite these criticisms, the Angle method of classifying malocclusion is considered to be the most practical (Moyers, 1988).



The Angle classification has been widely used in assessing the prevalence of malocclusion in communities. (Goose et al., 1957; Walther, 1960; Miller and Hobson, 1961; Ast et al., 1962; Heffer and Lovius, 1963; Moss and Picton, 1968 and Murray, 1968).

Ackerman and Proffit System

Ackerman and Proffit (1969) proposed a system of classification, based on a minimum of five characteristics, which they felt should be considered and systematically described. This approach was designed to overcome the weaknesses of Angle's classification. This system of classification is a

synthesis of both the Angle classification and the five identified characteristics of malocclusion, within a Venn diagram¹

The Venn diagram (Figure 1) consists of an evaluation of facial proportions and aesthetics, alignment and symmetry within the dental

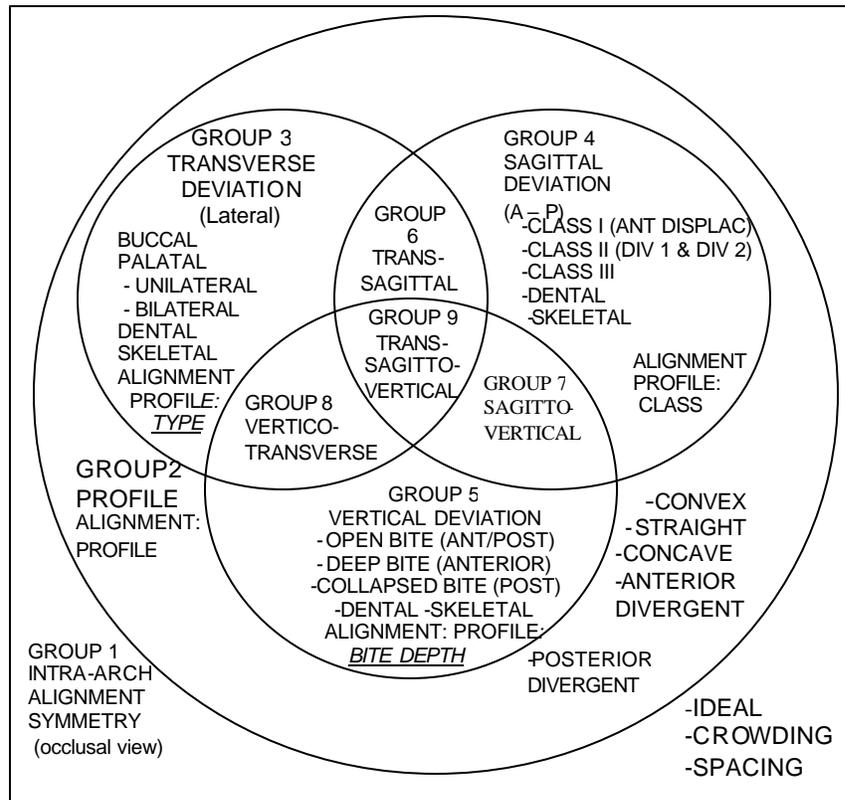


Figure 1. Ackerman and Proffit System

arches and skeletal and dental relations in the transverse, anteroposterior and vertical planes of space. It is grouped as followings:

¹ A Venn diagram offers a visual demonstration of interaction or overlap among parts of a complex structure. A collection or group in this system is defined as a set, and all elements contained in a set have some common property.

Group 1 - represents the universe

Group 2 - in this group, the profile is represented as a major set within the universe. Lateral (transverse), anteroposterior (sagittal), and vertical deviations and their interrelationships.

Groups 3 to 9 - are represented by three interlocking subsets within the profile set. Thus group 9 will represent the most complex malocclusion.

The complexity of this classification has limited its widespread application (Moyers, 1988).

EPIDEMIOLOGICAL INDICES

These indices are useful tools in determining the prevalence of occlusal anomalies in populations and are also valuable for research and human resource management. The most important requirement of any index used in this way is that it be reliable (WHO, 1966).

Index of Tooth Position

The Index of Tooth Position was proposed by Massler and Frankel (1951) as a quantitative method of evaluating malocclusion for epidemiological purposes. This method of assessment is based on the identification of individual teeth as units of occlusion rather than arch segments. Tooth displacement, rotation, infra-occlusion and supra-occlusion are recorded and the number of maloccluded teeth summed to give an overall measure of malocclusion.

The index is not reliable because of the difficulties encountered in judging the conformity of each tooth to an ideal position in all planes of space (Otuyemi and Jones, 1995). Furthermore, because each tooth is recorded as either maloccluded or aligned, a mildly displaced tooth scored the same as a severely displaced tooth, thereby giving no indication of the degree of severity.

The Dentofacial Index

In 1953 Elsasser developed the Dentofacial Index (DFI) as an epidemiological tool. It measures the dentofacial morphology using facial landmarks and certain features of malocclusion, that is, the presence or absence of crossbite and crowded dental arches. The facial orthometer (Figure 2), which is a dedicated instrument, was developed to facilitate measurement. This instrument assesses dentofacial pattern. However, this index has been found to be of greater value in anthropological studies.

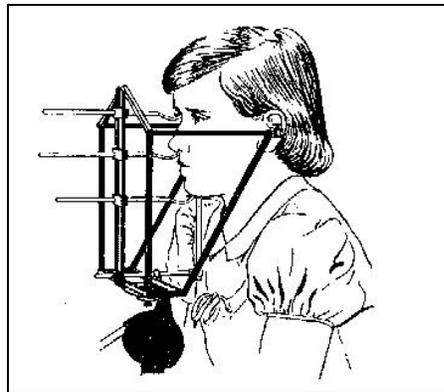


Figure 2. Child in position ready for orthometric measurement (Elsasser, 1953)

The Malalignment Index

The Malalignment Index (MI) was developed by Van Kirk and Pennell (1959) to assess malocclusion in population groups. The dentition is divided into segments that are assessed in the following order: maxillary anterior, right and left posterior and the mandibular anterior, right and left posterior. For each tooth present, two traits are considered namely; rotation and displacement. These are measured by means of a small plastic gauge specifically designed for this index. A score of 0 is awarded for ideal alignment, 1 for minor malalignment and 2 for major malalignment. The values are then summated to give a full-mouth index. They claimed the examination procedure to be rapid and simple.

Otuyemi and Jones (1995) found that this method of scoring does not reflect the true severity of malocclusion because the relationship of the teeth in occlusion is not taken into account.

The Occlusal Feature Index

The Occlusal Feature Index (OFI) was developed at the National Institute of Dental Research in 1957 (Poulton and Aaronson, 1961). This index is based on four primary features of occlusion which are considered to be of importance in an orthodontic examination. These include the following; lower anterior crowding, cuspal interdigitation, vertical overbite and horizontal overjet. Each of these four categories is scored, the total of which indicates the severity of the malocclusion. The totalled scores range from 0 to 9, with zero denoting 'normal' occlusion.

In a preliminary test of the OFI, Poulton and Aaronson (1961) claimed the index to have a reasonable inter-examiner reliability and good correlation with treatment need.

Tang and Wei (1993) considered this index incomplete since only four features of occlusion were measured and scored.

Björks' Method

Björk, Krebs and Solow (1964) introduced a method of recording malocclusion for epidemiological purposes. This consisted of a systematic registration of carefully defined "individual symptoms" (anomalies and/or deviations) based on three main features:

- 1) Anomalies in the dentition, that is; tooth anomalies (supernumery teeth, aplasia, malformation), abnormal eruption and malalignment of individual teeth.
- 2) Occlusal anomalies, that is; deviations in the positional relationships between the upper and lower dental arches.
- 3) Deviation in space conditions, that is; spacing or crowding of the teeth.

A specially designed instrument is used in the measurement of mandibular overjet, openbite, spacing, transverse forced bite, displacement of the midline and medial diastemas. The index also includes a subjective assessment of treatment need.

The registration is rather complex and simplification of the examination procedure is necessary as the total number of 567 features are recorded on the score sheet (Otuyemi and Jones, 1995).

Method For Measuring Occlusal Traits

A method for measuring occlusal traits was developed by the Working Group 2 of the Fédération Dentaire Internationale (FDI) Commission on Classification and Statistics for Oral Conditions (COCSTOC) (Baume *et al.*, 1973) which was to provide investigators and health authorities with a common basis for assessing the prevalence of malocclusion in various parts of the world. This simple, objective method for measuring occlusal traits was developed and field tested in the period 1969 to 1978.



The system sets out to measure and record, in a simple manner, three categories of occlusal features namely; dental, intra-arch and inter-arch relationships. Designated traits are recorded using codings for aspects of malocclusion, together with the FDI system of tooth identification to localise individual tooth malrelations (Baume *et al.*, 1973).

This examination should not be made on subjects who are still in the mixed dentition stage of development, because many occlusal problems in that stage of development are self-correcting (Bezroukov *et al.*, 1979).

A Quantitative Assessment Of Occlusal Features

Kinaan and Burke (1981) proposed a simple, reproducible and quantitative method of assessing occlusion for epidemiological studies. Five main features of occlusion were considered namely overjet, overbite, posterior crossbite, buccal segment crowding and incisor segment alignment. Each arch was divided into three segments, one anterior and two posterior, and these were assessed in terms of alignment and inter-arch relations. Four instruments were developed to facilitate direct intra-oral measurements, namely the depth gauge, overbite gauge, modified vernier calipgauge and a modified dial calipgauge. The need for these instruments when applying the index poses a shortcoming for the index.



INDICES OF TREATMENT NEED

Several Indices have been developed to attempt to categorise malocclusion into groups according to the level of treatment need (Shaw, 1991). These indices are valuable when allocating limited resources to priority groups or as guides in an orthodontic risk/benefit analysis.

The Handicapping Labio-Lingual Deviation Index

The Handicapping Labio-lingual Deviation (HLD) Index was developed by Harry L. Draker (1960) to determine the presence or absence of a physical dento-facial handicap for public health purposes. The social acceptance of individuals, in school or in the workplace, is often influenced by their physical appearance. Draker proposed that the factors causing

disfigurement may be based on seven components; cleft palate, traumatic deviations, overjet, overbite, mandibular protrusion, open bite, labiolingual spread which could be measured by the HLD Index. The index is applicable only to the permanent dentition.

An advantage of this index is that only a Boley gauge is required for measurement; special equipment is unnecessary. Also the scores permit differentiation between handicapping and non-handicapping malocclusions (Otuyemi and Jones, 1995). This index can be applied to both patients and models.

Carlos and Ast (1966) tested the ability of the HLD Index to distinguish handicapping from non-handicapping malocclusions. Clinical judgement made by orthodontists was used as the standard. The distributions of the HLD Index scores in the two groups were found to be largely overlapping, indicating an inability of the index to identify the so called handicapping malocclusion.

In their study Han and Davidson (2001) found that the index failed to identify:

- a) localized crowding that significantly compromises dental aesthetics (for example, rotation of maxillary central incisor),
- b) missing teeth or spacing in the anterior dental segment,
- c) asymmetry and

d) dysfunctional components such as posterior open bite, speech difficulties, and symptoms of temporomandibular joint dysfunction.

Maryland and California Modifications

In 1985, the American Association of Orthodontics (AAO) formally rescinded its decision concerning the recognition of the Salzmann index as the national orthodontic health index. The AAO stated that it does “not recognize any index rating classification or coding system as a scientifically valid measure of the need for orthodontic treatment” (AAO Bulletin, 1990).

Subsequently, public health planners in fifteen states had adopted several occlusal indices. However, cut-off scores to determine eligibility for orthodontic treatment in the public sector was proposed in an arbitrary manner (Han and Davidson, 2001). As a result several states introduced modifications to the HLD index. The index proposed by the state of Maryland (Md) suggested that the cut-off score of the HLD index be increased by modifying Draker’s scoring formula by subtracting 2mm from overjet and 3mm from overbite measurements (Han and Davidson, 2001).

In 1998, the HLD index was also modified in California to identify the most severe malocclusions. The HLD (CalMod) index included deep impinging bites and crossbites of individual anterior teeth with tissue destruction. Later, overjets greater than 9 mm, reverse overjets of 3.5 mm and unilateral posterior crossbite were also added as weighted factors. Thus a

more comprehensive index was developed that incorporated both aesthetic and functional components of malocclusion (Han and Davidson, 2001).

In a comparative study, of the HLD (CalMod) and the HLD (Md), Han and Davidson (2001) found that the correlation between the two indices was not very strong. In the HLD (Md) index crowding was a major determinant whereas in the HLD (CalMod) overjet was. The results also showed that the HLD (CalMod) index identified more patients with severe Class II malocclusion, for treatment, than did the HLD (Md).

Malocclusion Severity Estimate

The Malocclusion Severity Estimate (MSE) was developed by Grainger (1961) at the Burlington Orthodontic Research Centre. It consists of seven weighted and defined measurements namely: overjet, overbite, anterior openbite, congenitally missing maxillary incisors, relationship of the first permanent molars, posterior crossbite and tooth displacement. Six malocclusion syndromes were derived from the preceding measurements:

1. Positive overjet and anterior open bite
2. Positive overjet and overbite, distal molar relationship and buccal posterior crossbite
3. Negative overjet, mesial molar relationship and posterior crossbite
4. Congenitally missing maxillary incisors
5. Tooth displacement
6. Potential tooth displacement.

The final MSE score was that of the syndrome with the largest value, regardless of the scores of the other syndromes.

Table 1. Levels of severity of a malocclusion as established by the Malocclusion Severity Estimate.

<i>Interpretation</i>	<i>Levels</i>
Virtually classic "normal occlusion"	0
Minor manifestations of malocclusion and treatment need is slight	1 – 3
Definite malocclusion but treatment elective	4 – 6
Severe handicap, treatment highly desirable	7 – 9
Very severe handicap with treatment mandatory	> 10

In the MSE the absence of occlusal disorders was not scored as zero (Ghafari, 1989).



Treatment Priority Index

Grainger (1967), in an attempt to improve on the MSE, developed the Treatment Priority Index (TPI). It was formulated from the evaluation of models or clinical examination of 375 twelve-year-old children from three Ontario communities. The TPI differed from the MSE in that it eliminated the category of potential tooth displacement (syndrome 6) and also by rating distocclusion and mesiocclusion equally (Ghafari, 1989). This index is based on the inter-relationships of ten manifestations of malocclusion namely; bimolar relationship, maxillary overjet, openbite, overbite, tooth displacement, congenitally missing teeth, unerupted central incisors,

mandibular prognathism and retrognathism, and posterior crossbite. An eleventh feature was included for gross dento-facial anomalies from these the index defines seven natural groupings that tend to occur jointly and are referred to as syndromes. Based on these finding five grades of treatment need were developed.

Turner (1990) conducted a study using the TPI on children in their tenth year. In the first part of his study the validity of the index was investigated by three orthodontists ranking 134 study models. In light of the results obtained the TPI was then modified. The clinical judgement of two other orthodontists was compared with the scores from the modified TPI, using another 121 study models. Five Community Dental Officers were also instructed on the use of the modified TPI. Low levels of inter-examiner reproducibility were obtained. Turner (1990) concluded that firstly, the index was inadequate in identifying orthodontic treatment need in children during the mixed dentition stage and secondly, personnel untrained in orthodontics found difficulty in using the TPI.

In his study, Ghafari (1989) found the TPI to be a valid epidemiologic indicator of malocclusion but that it did not predict the severity of individual malocclusions in the permanent dentitions.

Handicapping Malocclusion Assessment Record

The Handicapping Malocclusion Assessment Record (HMAR) was developed by Salzman (1968) in response to a recommendation, in

1966, by the American Dental Association's Council on Dental Health. This index was designed to identify any occlusal condition that interferes with oral health or general well being, rather than for the identification and differentiation of various specific occlusal deviations. The total score in this index is made up of sub-scores reflecting intra- and inter-arch deviations within the anterior and posterior segments of both jaws. Intra-arch deviation is indicated by the number of teeth which are missing, crowded, rotated, or spaced; while inter-arch deviation refers to overjet, overbite, crossbite and anteroposterior relationships of the buccal segments. A third aspect includes the clinical assessment of six handicapping dentofacial deformities. Weightings are assigned to these deviations; which reflects estimates based on clinical experience with regard to problems of dental function, health and appearance. Using this index, 0 would indicate an ideal occlusal condition whereas 20 or more would indicate a severe handicapping occlusal condition (Abino, Lewis and Slakter, 1978).

That the HMAR is expeditious in use and requires no measurements. These were considered by Hermanson and Grewe (1970) to be its most important feature. Tang and Wei (1993) found that it records and weighs functional problems, which no other index does. Otuyemi and Noar (1996) found the HMAR to be simple, easy to use and widely accepted.

The HMAR has been criticised in that no matter how objective the recordings of the traits are, the weighting of the various occlusal characteristics is subjective (Brooke and Shaw, 1989; Shaw, 1991).

Another disadvantage of this index is that it does not include an aesthetic or psychological component, considered to be an important feature of an effective index by the American Association of Orthodontists Orthodontic Indices Consensus Conference in 1993 (Lindauer et al., 1998).

Occlusal Index

The variations in terminology, concepts and methodology was what motivated Summers (1971) to develop the Occlusal Index (OI). This Index is based on the Malocclusion Severity Estimate of Grainger (1961) and is an attempt to remedy its shortcomings. Nine other characteristics are included namely: tooth displacement, molar retention, overbite, overjet, posterior crossbite, posterior open bite, midline relations, missing permanent maxillary incisors and dental age.



The OI was shown to have the highest validity over periods of time (Summers, 1971; Gray and Demirjian, 1977). Grewe and Hagan (1972) proposed that it exhibited the least amount of bias; later confirmed by Tang and Wei (1993). Summers (1972) found the index to be best correlated with clinical standards. Pickering and Vig (1976) used the Occlusal Index for the assessment of treatment standards and felt that it was the most suitable index available for the assessment of treatment outcome. The OI is the only index that developed different scoring modalities for patients in different stages of dental development (Tang and Wei, 1993).

McLain and Proffit (1985), in their review of the prevalence of malocclusion, reported that despite these attributes, the Occlusal Index was not commonly employed worldwide.

Pickering and Vig (1976) found that the index was unable to cope with, and excluded, cases in which there had been loss of first permanent molars. Elderton and Clarke (1984) felt that some of the diagnostic criteria were not adequately defined. So and Tang (1993) criticized the Occlusal Index for failing to score or record spacing except in cases involving an upper median diastema greater than 2 mm. Otuyemi and Noar (1996) found the Occlusal Index to be time-consuming and cumbersome to use, involving a long, complex procedure of scoring, thereby making research and audit difficult. In addition, they also found that the index does not take into account buccal crossbites, openbites, centreline discrepancies or deep overbites that impinge on the lower labial or palatal gingivae.

Despite the criticisms of the Occlusal Index, it has been shown to be one of the most reliable and valid indices of treatment need (So and Tang, 1993).

The Swedish System

A priority index of need for orthodontic treatment was formulated, in 1966, by the orthodontic division of the Swedish Dental Society, and the Swedish Medical Board (Linder-Aronson, 1974). This system concentrated

on dental health impairment and proposed guidelines for measurement of aesthetic impairment. It is comprised of a four grade index scale:

Grade 1. Little need.

Mild deviations from normal (ideal) occlusion, for example; prenatal occlusions with little negative overjet, postnormal occlusion without other anomalies, deep bite without gingival contact, open bite with little frontal opening, cross-bite without posturing, mild crowding or spacing, inversion of single teeth without forced bite, mild rotations of only little cosmetic and/or functional significance.

It includes malocclusions that should be disregarded and anomalies in this group are not meant to be referred to a specialist.

Grade 2.

Aesthetic and/or functionally disturbing proclined or retroclined incisors, deep bite with gingival contact but without gingival irritation, severe crowding or spacing of teeth, infraocclusion of deciduous molars and permanent teeth, moderate frontal rotations.

Grade 3.

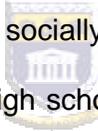
Postured bite, deep bite with gingival irritation, extreme open bites, cross-bite causing transverse forced bite, scissor bite interfering with articulation, severe frontal crowding or spacing, retained canines, cosmetically and/or functionally disturbing rotations.

Grade 4. Very urgent need.

Cosmetic and/or functionally handicapping anomalies, for example; cleft lip and palate, extreme post- and prenatal occlusion, retained upper incisors, extensive aplasia (Linder-Aronson, 1974).

The criteria for assessment were not well defined and the cut-off points were vague (Shaw et al. 1991). Additional criteria were added, by Ingervall and Ronnerman (1975), which incorporated a morphological index and a functional index.

The Dental Aesthetic Index

The Dental Aesthetic Index (DAI) developed by Cons et al. (1989) is an orthodontic index based on  socially defined aesthetic standards. Approximately 1,600 American high school students and adults rated 200 stimuli. These comprised of photographs of teeth in occlusion representing the full range of occlusal conditions found in a population of half a million people. The teeth portrayed in each photograph were the incisors, the canines, and the first and second premolars in both the maxilla and mandible, which were framed by stylised masks in the shape of lips. These photographs were completely neutral with regard to sex, race or ethnic origin. There were no confounding facial features that could influence subjects' assessments of the social acceptability of the stimuli. A unique feature of the occlusal conditions depicted in each stimulus was the availability of 49 occlusal trait measurements.

Mean scores for social acceptability of the 200 stimuli, rated by the public, were linked by factor analysis and stepwise regression procedures to the occlusal trait measurements that were available from each of the photographs. The resulting regression equation consisting of ten components (intra-oral measurements of occlusal traits) and their appropriate regression coefficients (weights) is referred to as the Standard Dental Aesthetic Index (Cons et al., 1989). The components and weights of the DAI are shown in Table 2 (Jenny et al., 1993).

Table 2. Components and weights of the Dental Aesthetic Index

Component	Weight
Constant	13
Missing incisor, canine and premolar teeth	6
Crowding in incisal segments (No. of segments)	1
Spacing in incisal segments (No. of segments)	1
Diastema – in millimetres	3
Largest anterior irregularity, maxilla - in millimetres	1
Largest anterior irregularity, mandible – in millimetres	1
Anterior maxillary overjet – in millimetres	2
Anterior mandibular overjet – in millimetres	4
Vertical anterior openbite – in millimetres	4
Antero-posterior molar relation – largest deviation from normal ($\frac{1}{2}$ cusp = 1, full cusp or more = 2)	3
Total	DAI score

The treatment need scores as determined by the DAI are:

25 and below - Normal or minor malocclusion; no treatment needed

26 to 30 – Definite malocclusion; elective treatment.

31 to 35 – Severe malocclusion; need treatment

36 and higher – Handicapping malocclusion; need treatment.

The score can be placed on a continuum to determine the point at which the individual falls between most and least aesthetically pleasing dental appearance. The further a DAI score deviates from the norm the more likely the occlusal condition may be judged as socially or physically handicapping (Jenny and Cons, 1996).

The index is useful in predicting handicapping and non-handicapping malocclusions (Jenny et al., 1993). It is used in epidemiological surveys, to identify the need for orthodontic treatment, and as a screening tool to determine priority of treatment (Ansai, 1993). Similarly to the Occlusal Index, the DAI has been used to assess treatment standards (Lobb, 1994). It has also been integrated into the items of the International Collaboration Study of Oral Health Outcomes (ICS II) by the WHO in 1989 (Beglin et al., 2001).

The ease of measurement of the DAI traits was noted by Kaey et al. (1993) who suggested that without a comprehensive orthodontic evaluation, it would substantially overestimate the number of individuals requiring treatment.

The Standardized Continuum of Aesthetic Need

The Standardized Continuum of Aesthetic Need (SCAN) index (Evans and Shaw, 1987) was based on the perception of dental aesthetics in the

United Kingdom. Individuals matched their dental appearance against ten photographs rated (unknown to them) from 1 (attractive) to 10 (unattractive). This is then used to record the aesthetic impairment of the malocclusion in the individual.

In a study by Flores-Mir et al. (2004), the subjects seemed to have difficulty understanding the idea of the selection of the appropriate photograph from the ten provided. Some tried to match the photographs most resembling their own teeth, instead of selecting one that had the same level of aesthetic appeal as their own. Burden, (1995) reported that professionals, during their initial training in the use of the scale, also experienced this problem.

Index of Orthodontic Need



The Index of Orthodontic Need (IOTN) developed by Brooke and Shaw (1989) is a combination of the Standardized Continuum of Aesthetic Need index (Evans and Shaw, 1987) and the Swedish System (Linder-Aronson, 1974). It ranks malocclusion in terms of the significance of various occlusal traits related to the individual's dental health and perceived aesthetic impairment, with the intention of identifying those individuals who would be most likely to benefit from orthodontic treatment. The index comprises an aesthetic and a dental health component (Shaw , 1991).

The Aesthetic Component

The Aesthetic component (AC), derived the SCAN index of Evans and Shaw (1998), consists of a 10-point scale illustrated by a series of numbered photographs (Appendix I). These photographs represent three categories: no treatment need (grades 1-4), borderline need (grades 5-7), and great treatment need (grades 8-10). A rating is allocated for overall dental attractiveness rather than specific morphological similarity to the photographs. The value arrived at gives an indication of the patient's treatment need on the grounds of aesthetic impairment, and by inference reflects the sociopsychological need for orthodontic treatment (Shaw et al., 1991).

Obvious shortcomings of the scale are its poor ability to represent dentofacial imbalance in the anteroposterior plane (Evans and Shaw, 1987). Buchanan et al. (1994) applied the IOTN clinically to a group of patients and later to their study models and photographs. A poor agreement was obtained for the AC scored from photographs, as compared with those scores recorded clinically and from the study models. The poor levels of agreement are due to the fact that photographs are two-dimensional representations of three-dimensional objects. Photographs reduce the prominence of anterior irregularities and overjet problems. Mattick et al. (2004) found that the diminutive photographs may bias the AC in a favourable (lower) direction by masking minor irregularities.

The Aesthetic Component takes account of the teeth only and not the teeth "within the face" (Evans and Shaw, 1987). McGuinness and

Stephens (1995) found that this may be a shortcoming of the IOTN when used in epidemiological investigations of Black subjects. De Mûelenaere et al. (1998) recommended that minor adjustments of these epidemiological instruments may be needed in order to make them more applicable to other population groups.

Dawjee et al. (2002) undertook a study among black evaluators to determine their perceptions. Their results indicated that features common to individuals with bimaxillary protrusion such as anterior open bite; anterior diastemas and reverse overjet (Trottman and Elsbach, 1996) were of no importance to the respondents. None of the final 5 selected photographs differed much from the original aesthetic component of the IOTN indicating that the concerns of black subjects were congruent to the findings of Evans and Shaw (1987).



The results of a study by Hlongwa et al. (2004) found that both the IOTN and the DAI are capable of assessing malocclusion severity in both Black and White subjects and further determine their orthodontic treatment needs.

The Dental Health Component

The Dental Health component (DHC) is loosely based on the Swedish System. Each occlusal trait thought to contribute to the longevity and the satisfactory functioning of the dentition is defined. With the use of a specially designed ruler, various features of the malocclusion can be

noted, measured and placed into five clearly defined grades, with clear cut-off points between the grades (Shaw et al., 1991). The DHC categorises cases from grade 1 (no need for treatment) to grade 5 (great need) and may be applied both clinically and to study casts.

A fundamental premise of the index is that it identifies dental diseases that are site specific, (for example severe displacement of a particular tooth represents a particular disadvantage for that site), and the most severe trait identified is the basis for grading the individual's need for treatment (Shaw et al., 1991).

The problem with using the Dental Health Component of the IOTN is that minor irregularities may not score high enough to place the patient in a treatment need category (McGuinness and Stephens, 1995).



TREATMENT OUTCOME INDICES

Orthodontists and health care providers have displayed increased interest in assessing the efficiency of orthodontic treatment for correction of malocclusion, however, this is difficult to quantify. To date, treatment need indices have been used to assess treatment outcomes (Elderton and Clarke, 1983 and 1984; Lobb et al., 1994; Richmond et al., 1994a; Richmond and O'Brien, 1996). None of the indices used these studies, namely the Occlusal Index (Summers, 1972), the Dental Aesthetic Index (Cons et al., 1986) and the IOTN (Shaw et al., 1991a) have been designed or validated for this purpose. At best these indices measure the degree of

residual treatment need, but this may not be sufficiently quantitative to assess significant differences in treatment efficacy.

The Peer Assessment Rating Index

The Peer Assessment Rating (PAR) Index was developed in the United Kingdom to record “malocclusion” at any stage of treatment (Richmond et al., 1992). This index was carefully tested for reliability and validity. It was developed over a series of meetings of a group of experienced orthodontists (British Orthodontic Standards Working Party, 1986). More than 200 study casts were examined and discussed until consensus was reached regarding individual features considered to be important in obtaining an estimate of malocclusion. A score was then allocated to each feature that deviated from the ideal, and component scores were added to obtain a total score representing the degree of malocclusion. A ruler was also developed to allow rapid analysis of study casts.

The index was validated using assessments of deviation from normal occlusion (the Gold standard). Attempts to improve validity were done by assigning multipliers or weightings to each component thus producing a new weighted PAR score. This was the final form in which the index was introduced (Richmond et al., 1992a).

The PAR Index is used to measure treatment outcome by comparing pre- and post-treatment scores for point and percentage reductions.

Improvement is categorized into three grades according to specific criteria; 'Greatly improved' requiring a score reduction greater than 22 points, 'Improved' requiring a reduction between 0 and 22 points, and 'Worse or no different' categorized no reduction of the pre-treatment. The criteria are graphically represented using the 'PAR nomogram' (Richmond et al., 1992a).

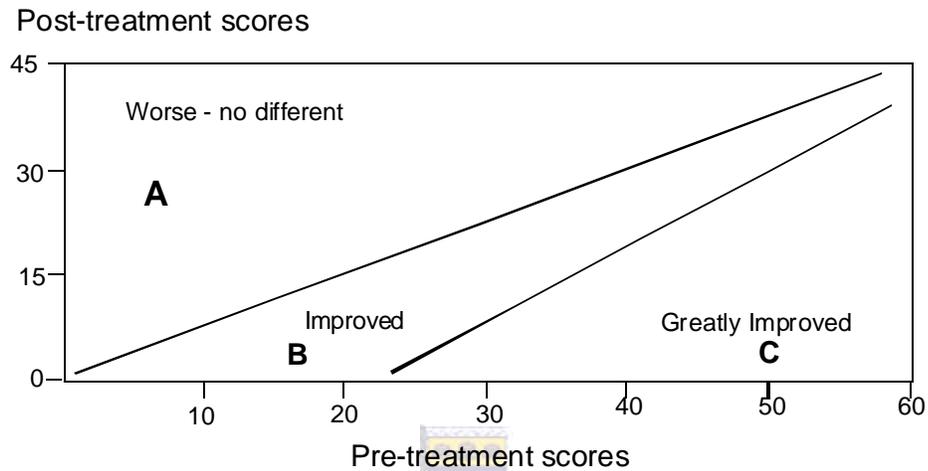


Figure 3. The Index of Treatment Standards nomogram

Currently, there are separate British and American weightings. The American weighting emphasises overbite, the buccal segments and the midline. The lower labial segment is excluded because it is not thought to influence the perception of treatment outcome. McKnight, et al. (1998), noting lower incisor relapse, argue that the lower labial segment alignment should be included in the US version. In contrast, the British version includes the lower labial segment and places a greater emphasis on overjet.

PAR weightings proposed for USA and UK		
	USA PAR weighting	UK PAR weighting
Overjet	4.5	6
Overbite	3	2
Midline	3.5	2
Buccal Occlusion	2	1
Upper labial segment	1	1
Lower labial segment	0	1

Excellent reliability within and between trained examiners has been demonstrated (Richmond et al., 1992; O'Brien, Shaw and Roberts, 1993; De Guzman et al., 1995). The PAR index has therefore gained a considerable measure of acceptance (Mcknight et al., 1998).

Studies have identified limitations associated with PAR scoring (Fox, 1993; Kerr et al., 1993). Problems relate mainly to the generic weighting system, particularly that for overjet and overbite. The index may be unduly sensitive to increased overjets, for example an overjet reduction from 8mm to 2mm, by retroclining the upper incisors, will reduce the PAR score by 18 points - only 4 points from 'Greatly improved' (according to the PAR nomogram), whereas the weighting for overbite is so low that the correction of a complete and traumatic overbite merits a reduction of only 6 points. Thus failing to represent treatment value in terms of function and appearance.

Shaw (1995) highlighted that the PAR cannot identify inappropriate expansion or incisor inclination, it also cannot measure improvements in

appearance or psychosocial well-being. As deciduous teeth are excluded from the PAR, it does not score highly for the mixed dentition.

A further limitation of the PAR Index is that occlusions with initial scores of less than 22 points cannot become 'Greatly Improved' after treatment. The zero weighting allocated to 'displacements' in the buccal segments, which include impacted teeth, indicate that such irregularities are disregarded even though their correction may have a significant effect on treatment outcome (Brooke and Shaw, 1989).

Turbill et al. (1996) highlighted the limitation of the UK weighting, suggesting that the buccal occlusion weighting is too low and the overjet weighting too high.



Hamden and Rock (1999) found difficulties arising in the application of only one weighting system for all malocclusions, since occlusal features vary in importance in different classes of malocclusion. Therefore, to establish the validity of the PAR Index they used the subjective judgements of orthodontists (as the Gold standard), clinical ranking of occlusal features and statistical modelling to derive a new weighting system, separate for each malocclusion class. As a result a new and more sensitive method of assessment was suggested which utilizes a combination of point and percentage reductions in PAR scores. The new weighting system was found to have better correlations with the Gold standard than the PAR nomogram.

Turbill et al. (1996) suggested that the PAR and IOTN would need to be updated from time to time to keep them abreast of current perceptions and knowledge within the profession.

INDICES OF TREATMENT COMPLEXITY

Stephens and Harradine (1988) reported that the proportion of patients receiving complex treatment at a UK dental teaching hospital, between 1977 and 1985, had increased greatly. To date, no specific index has been developed to measure only treatment complexity, although such an index would be useful in setting fee levels objectively, particularly in a State-funded structure.

COMPARATIVE STUDIES



Hermanson and Grewe (1970) compared several indices, including the Malocclusion Severity Assessment, the Treatment Priority and the Occlusal Index to each other. Of these the Occlusal Index was found to be the most objective and easiest to use, although the most difficult to learn. Clinicians not familiar with the Occlusal Index showed a high degree of agreement and consistency between their assessment and the Occlusal Index scores. The Treatment Priority Index and the Occlusal Index showed the best precision and the least bias.

Grewe and Hagan (1972) compared the Handicapping Malocclusion Assessment Record, the Occlusal Index and the Treatment Priority Index.

They concluded that none of the indices showed a significant difference in precision or examiner variability. But the Occlusal Index was given as the index of choice because it demonstrated the least amount of bias.

Gray and Demirjian (1977) compared the reproducibility and accuracy of four indices: the Handicapping Labiolingual Deviation Index, the Treatment Priority Index, the Occlusal Index and the Handicapping Malocclusion Assessment Record. The results showed that all methods were highly reproducible, but that the OI had the best correlation with the clinical standard.

Järvinen and Väättäjä (1987) examined the variation in the results of four somewhat different treatment need indices when measuring the severity of malocclusion and the need for orthodontic treatment, namely the Swedish System, HMAR, TPI and the Index for Need of Orthodontic Treatment (INOT) of Ingervall and Rönnerman, (1975),. The study showed a marked variation between the indices, and indicated that the different methods selected different groups of children needing treatment.

Otuyemi and Noar (1996) set out to assess the variation in time spent in recording malocclusions using the HMAR, OI and DAI and to determine the relationships that exist between them. Their results showed that HMAR and OI took a significantly longer time to execute than the DAI, and that the OI was the most time consuming. All the indices correlated extremely well with one another. Excellent levels of reliability were

achieved with all the indices with the DAI demonstrating the highest level of reliability.

A comparative study using both the Occlusal Index and the Index of Orthodontic Treatment Need to assess a group of 100 dental students was undertaken by So and Tang (1993). They found that both indices were highly reproducible, but there were significant discrepancies in the treatment needs as assessed by the two indices. Furthermore their findings showed that the IOTN tended to overestimate, while the OI tended to underestimate treatment need where there were missing or extracted teeth. The IOTN appeared to overestimate 'tooth displacement' and 'crossbites'; while the OI did not score missing teeth, except in cases of missing maxillary permanent incisors. The OI also did not score mesiodistal or buccolingual tipping of teeth that would occur subsequent to tooth loss, thus contributing to its underestimation.

Beglin et al. (2001) compared the reliability and validity of the Dental Aesthetic Index, the Handicapping Labiolingual Deviation Index (CdMod) and the Index of Orthodontic Treatment Need. They concluded that the three indices are reliable and valid instruments with which to determine treatment need.

Freer and Freer (1999) compared the screening methods of the Dental Aesthetic Index, the Index of Orthodontic Treatment Need and the Danish Ministry of Health (Solow, 1990). This study highlights the differences in

recommendations for treatment measures by the indices as opposed to the judgement of the orthodontist. According to them a decision for or against treatment of the individual based on the index score only will always be open to challenge.

THE INDEX OF COMPLEXITY, OUTCOME AND NEED

More recently, Daniels and Richmond (2000) developed the Index Of Complexity, Outcome And Need (ICON). They felt that the same measurement tool used to assess treatment need should be used to assess treatment outcome. The index is intended for the use in the context of a specialist practice, to provide a means to compare treatment thresholds in different countries and to serve as a basis for quality assurance standards in orthodontics.

An international panel of 97 orthodontists from nine countries judged a sample of 240 dental casts for the assessment of treatment need and further 98 paired pretreatment and post-treatment cases for assessment of treatment outcome. The practitioners each gave a dichotomous decision on the need for treatment and the acceptability of the treatment outcome. Furthermore, the practitioners gave a judgement (using 5-point rating scales), for the pretreatment complexity and post-treatment degree of improvement. The mean complexity and improvement rating was then calculated for each case. The authors examined the dental casts, and occlusal traits in the sample were then comprehensively scored according to an objective scoring protocol (Richmond and Daniels, 1998a). The occlusal traits scored included:

- (1) upper and lower labial segment alignment;
- (2) anterior vertical relationship, centreline, impacted teeth, upper and lower buccal segment alignment (left and right added together), buccal segment antero-posterior relationship (left and right added together), buccal segment vertical relationship (left and right added together), crossbite, missing teeth for any reason (excluding third molar);
- (3) aesthetic assessment based on IOTN aesthetic component, overjet in millimetres (centred at 3 mm), reverse overjet in millimetres, upper and lower incisor inclination relative to the occlusal plane, overall upper arch crowding/ spacing, overall lower arch crowding/spacing, lip competency.

The practitioners' subjective judgements of the casts were then related to the occlusal trait scores for each case using regression analyses. The Stepwise Multiple Logistic Regression was used to identify occlusal traits which were useful to predict the practitioners yes/no decisions (treatment versus no treatment and accept outcome versus reject outcome). Initially, separate predictive equations were calculated for treatment need and outcome decisions. Fortuitously, the equations for the two decisions identified similar (though not identical) occlusal traits. This finding led to the use of a set of five occlusal traits (identified in the initial analyses) to predict both dichotomous decisions. Initially, weightings for the five occlusal traits were calculated for the treatment need and outcome decisions separately, then a single set of weightings was tested which was based on the average of the two weightings for each occlusal trait. These are all shown in Table 3. The set of 'average' weightings formed a new

single index of treatment need and outcome assessment (Daniels and Richmond, 2000).

Table 3. Logistic regression weightings for treatment need and outcome

Logistic regression weightings for treatment need, treatment outcome, and the combined index using occlusal trait scores as explanatory variables.			
Term	Weightings for Treatment need	Weightings for treatment outcome	'Average' index weighting
IOTN Aesthetic Component	0.8420	0.5914	0.7
Left + Right buccal anteroposterior	0.3032	0.3030	0.3
Upper arch crowding	0.6036	0.2519	0.5
Overbite/open bite	0.4927	0.3876	0.4
Crossbite	0.6460	0.5091	0.5

According to Daniels and Richmond (2000), the new index is relatively simple to use requiring, no hierarchy (in reference to the IOTN) and having relatively few traits to measure. Most of the measurement protocols are common to components of PAR or IOTN, so there is already experience in the use and teaching of most of the occlusal traits. Application of the index takes approximately 1 minute for each case and, therefore, it is relatively quick. It requires no measurement tools other than an ordinary millimetric rule and an Aesthetic Component scale (Shaw *et al.*, 1991a). The index is intended for use in the late mixed dentition onwards.

Fox and his co-workers (2002) evaluated whether there was any relationship between the ICON, IOTN and PAR. They also wanted to establish whether or not the ICON could replace these indices as a

measure of orthodontic treatment complexity, outcome and need. Their findings showed significant correlations between IOTN and ICON with respect to treatment need, and between the PAR and ICON with respect to outcome. The authors suggested that this single index can replace the PAR index and the IOTN.



**CHAPTER THREE: AIM, OBJECTIVES AND
METHODOLOGY**



AIM

To assess the extent to which the ICON fulfils the requirements of an ideal index.

OBJECTIVES

The objectives of the study were to:

1. Establish a gold standard against which the reliability and validity of the components of the ICON could be determined.
2. Assess ease of use and simplicity of the ICON.
3. Assess the applicability of the ICON on patients and corresponding study models.



METHODOLOGY

Study Design

This was a quantitative study testing whereby the ICON was tested against an ideal.

The Sample

A sample of a 125 pre- and post-treatment study casts were obtained from the Orthodontic Department at the University of the Western Cape.

An additional sample comprising 60 untreated patients was identified from those examined for treatment in the Orthodontic department.

Data analysis

The data was captured on Microsoft Excel worksheets and analysed using the Microsoft Analyse It.

Materials

The materials used included the following:

- Pre- and post treatment study models
- ICON scoring protocol
- Aesthetic scale of the IOTN
- A millimetre ruler

Method

The study was performed in three parts to achieve the following objectives:



Objective One.

To Test The Validity And Reliability Of The ICON

Five Orthodontic specialist (experts), with an average of 25 years (range: 12 to 42 years) experience, were invited to rate pre- and posttreatment study models (n = 100). These models represented a range of occlusal conditions varying from mild to severe malocclusions, as determined by the IOTN. The unranked study models were displayed in numerical order on bench tops in a large room.

The panel of experts was required to assess the Treatment Need, Treatment Outcome, Complexity, and Degree of Improvement for each set of study models by means of a questionnaire (Appendix II). They were instructed to rate the models as objectively as they possibly could. No specific definitions were given for any of the ICON components.

Each of the 100 study casts was to be assigned, by the raters, to a “treatment” or “no treatment” category which was compared to the ICON score obtained by the researcher. The developers of the ICON had proposed that the cut-off point for treatment need be a score > 42 (Daniels and Richmond, 2000).

For treatment outcome each orthodontist was asked to decide whether the treatment was acceptable or not acceptable using their clinical judgement.



In addition, they were asked to assess the complexity in each case by using the following 5-point scale:

1. easy
2. mild
3. moderate
4. difficult
5. very difficult

Similarly for degree of improvement the following scale was used:

1. greatly improved,
2. substantially improved,

3. moderately improved,
4. minimally improved, and
5. not improved or worse.

The raters scored the study models at their own pace with no restriction on time.

A numerical value was assigned to each decision the raters made, for example, 'no treatment' = 0; or 'treatment need' = 1. To reflect the opinions of the experts, the view of the majority of the raters determined the Gold Standard. Therefore, if more than 2 raters gave a score of 1 the Gold Standard was set to 1, otherwise it was set to 0.

The researcher scored the 100 study models using the ICON and the Aesthetic Component of the IOTN. The scores obtained by the examiner were then compared to the gold standard to test the level of agreement.

A week later 25% of the study models was re-scored by the researcher to assess intra-examiner reliability.

Objective Two.

Assessing ease of use and simplicity of the ICON

The second part compared the assessments of the examiner with 4 orthodontic specialists and 3 non-specialists (2 dentists and an oral hygienist), who had not used the ICON previously.

Twenty-five other pre- and post-treatment study models were selected. A brief instruction on how to use the ICON score sheet was given. The time taken to score the study models was also noted.

Objective Three

Testing the ICON clinically and on study models

In the third stage a sample of 52 pre-treatment cases were obtained from the orthodontic waiting list at the University of the Western Cape. Each patient and corresponding study model was identified by means of a file number. All the patients were scored first; the matching study models were then identified and scored at a later stage.

Statistical Analysis

The simple Kappa test, which is a measure of agreement that has been corrected for the chance agreement, was used to assess the agreement of the index with the expert panel for treatment need and outcome. Weighted kappa statistics were used to assess both intra-examiner and inter-rater reliability, for complexity and degree of improvement. The scale of Fleiss

(1981) was used to interpret the strength of agreement for the kappa scores, (Table 4).

Table 4. The suggested subdivisions of Cohen’s Kappa statistic.

Less than 0.4	Poor agreement
Between 0.4 and 0.75	Fair to good agreement
Greater than 0.75	Excellent agreement

The method for calculating validity of the index is shown in Appendix III.

Ethical Statement

All the participants in the study were informed of the purpose of this study and of their right to refuse participation. Confidentiality was assured in every case. Verbal consent was obtained. All participants are currently undergoing treatment at the Faculty.



CHAPTER FOUR: RESULTS



This study was performed in three parts to: establish a gold standard against which the reliability and validity of the ICON would be determined, the ease of use and simplicity, to test the applicability of the ICON on patients and study models.

The statistical method used to measure reliability was Cohen's Kappa (Cohen, 1960), which assesses agreement between raters by eliminating chance. Fleiss (1981) suggested the subdivision of the kappa into ranges (Table 4); these have been used to interpret the results.

The IOTN was used on the sample (n = 100) to determine the distribution of the severity of the cases (Table 5) prior to assessment by the panel in order to ensure a wide variety of cases were included in the sample. The individual cases were classified by the IOTN in Table 5.

Table 5. Distribution of IOTN (DHC) grades in sample (n = 100)

IOTN grade	Treatment Need	Frequency
1	No Need	6
2	Little need	29
3	Borderline	14
4	Great need	40
5	Very great need	11
Total cases		100

The IOTN scores also served as an additional reference in determining treatment need.

Objective One

In the first part of the study, pre- and post-treatment study models (n = 100) were examined by five experts from which the gold standard was established. Thereafter, the examiner scored the same sample using the ICON and the Aesthetic Component of the IOTN to test for reliability and validity.

Inter-rater Reliability

The Inter-rater reliability was measured using the questionnaire shown in appendix II, with regards to *treatment need* and *outcome*, showed perfect agreement.

Table 6. Inter-rater agreement of the Experts with the Gold Standard

Components	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Complexity	0.48**	0.61**	0.70**	0.49**	0.56**
Degree of Improvement	0.62**	0.51**	0.57**	0.74**	0.57**

** fair to good agreement

For *Complexity* and *degree of improvement* the kappa showed an agreement of 'fair to good' (Table 6).

Intra-Examiner Reliability

Intra-examiner reliability was determined by re -scoring 25% of the sample. The kappa (κ) values indicated 'excellent agreement' for the categories of *treatment outcome* (100%), *complexity* ($\kappa = 0.80$) and *degree of improvement* ($\kappa = 0.84$). Agreement for *treatment need* ($\kappa = 0.63$) was 'fair to good' (Table 7).

Table 7. Intra-examiner Reliability (n = 25)

ICON COMPONENTS	KAPPA
Treatment Need	0.63**
Outcome	100% agreement***
Complexity	0.80***
Degree of Improvement	0.84***

** fair to good agreement

*** excellent agreement

Treatment Need

The ICON scored 25% of the cases as not needing treatment as compared to the gold standard, which determined that none of the cases fell into the 'no treatment' category (figure 4). However, the IOTN, scored 6% of the cases as 'no treatment need' (Table 5).

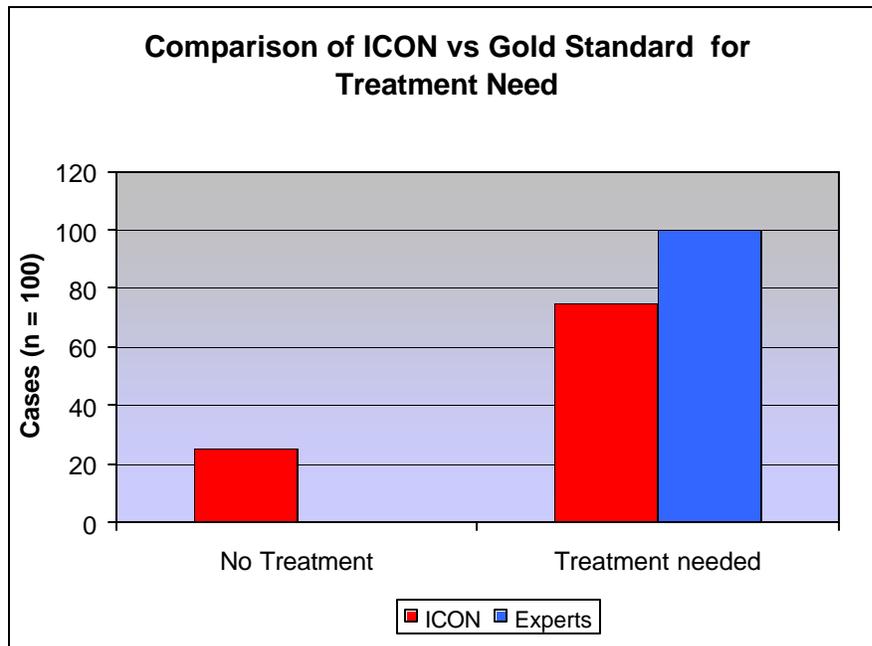


Figure 4. Treatment Need

Validity

The test for **sensitivity** (Appendix III) yielded a result of 75%.

Specificity was undefined as there were no negative values.

The **positive predictive value**, that is the number of cases that the index identified as needing treatment that in fact truly needed treatment, was 100%.

The experts scored no negative values; therefore the **negative predictive value** was 0%.

The **accuracy** of the test, which is the proportion of subjects that are correctly classified, was 75%.

Reliability

Reliability of the *treatment need* component was 'fair to good' ($\kappa = 0.63$).

Complexity

As illustrated in the frequency graph (Figure 5) there was a tendency for the raters to score predominantly in the 'moderate' category of the complexity scale. The experts assessed none of the cases as being easy.

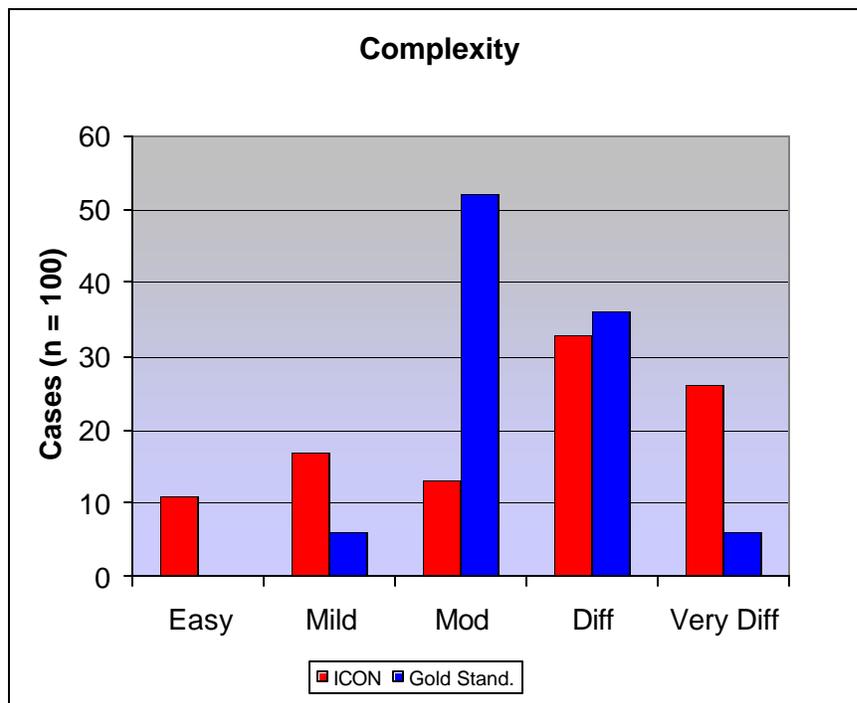


Figure 5. Complexity

Validity

When validity of the *complexity* component of the ICON was tested, the kappa revealed a 'poor agreement' value ($\kappa = 0.2$).

Reliability

Reliability for this component showed excellent agreement ($\kappa = 0.80$).

Treatment Outcome

There was excellent agreement (97%) between the gold standard and the index with regard to *treatment outcome* (figure 6).

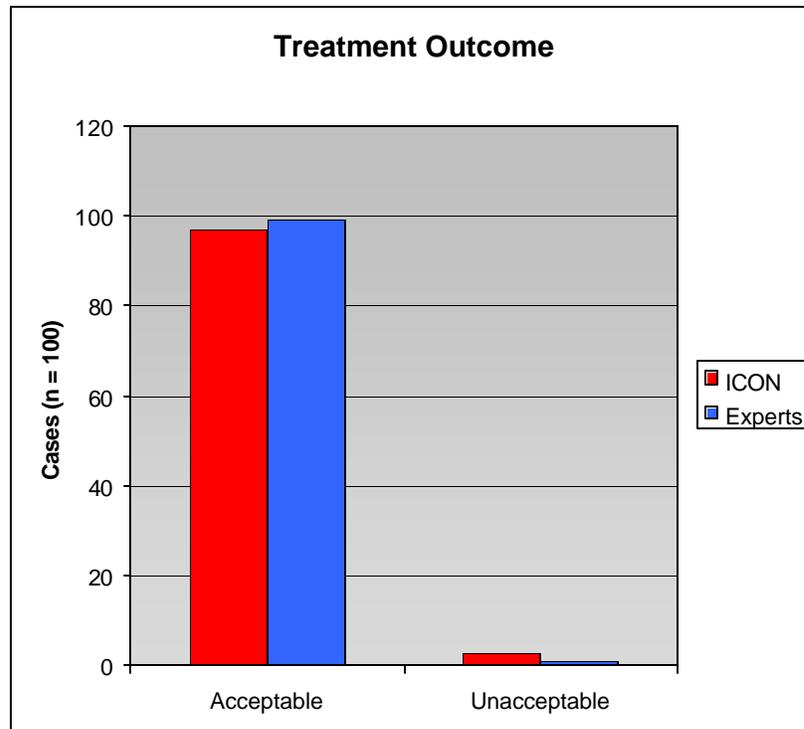


Figure 6. Treatment Outcome

Validity

The test for **accuracy** was 96% and **sensitivity** (Appendix III) yielded a result of 97%.

Specificity was calculated to be 0%.

The **positive predictive value** was shown to be 99%.

The **negative predictive value** was 0%.

Reliability

The reliability for *treatment outcome* showed a perfect agreement of 100%.

Degree of improvement

The distribution of the scores for the *degree of improvement* (figure 7) exhibits a large variation between the ICON and the gold standard.

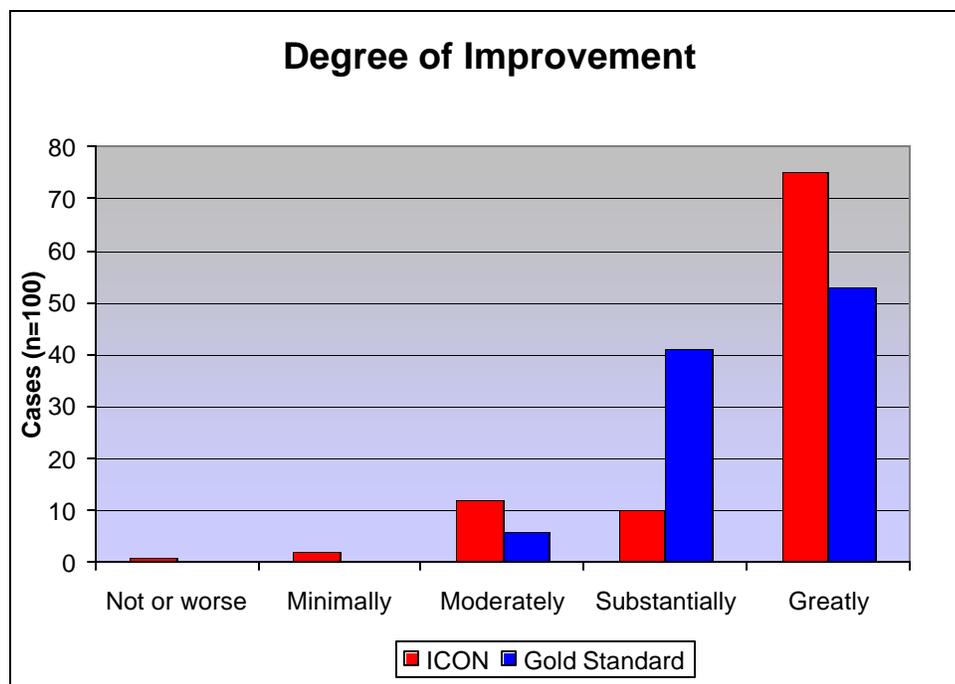


Figure 7. Degree of Improvement

Validity

Poor agreement ($\kappa = 0.34$) was found for *degree of improvement*

Reliability

The agreement calculated for *degree of improvement* was 'excellent' ($\kappa = 0.84$).

In summary, the values showed that the index is reliable with consistent application of the ICON over time. When the ICON was compared to the gold standard, validity for *treatment outcome* was 'excellent', for *treatment need* it was 'fair to good' and for *complexity* and *degree of improvement* it was 'poor'.



ANALYSIS OF THE “NO TREATMENT” GROUP

Further analysis was done to assess the distribution of the cases that the ICON identified as not needing treatment.

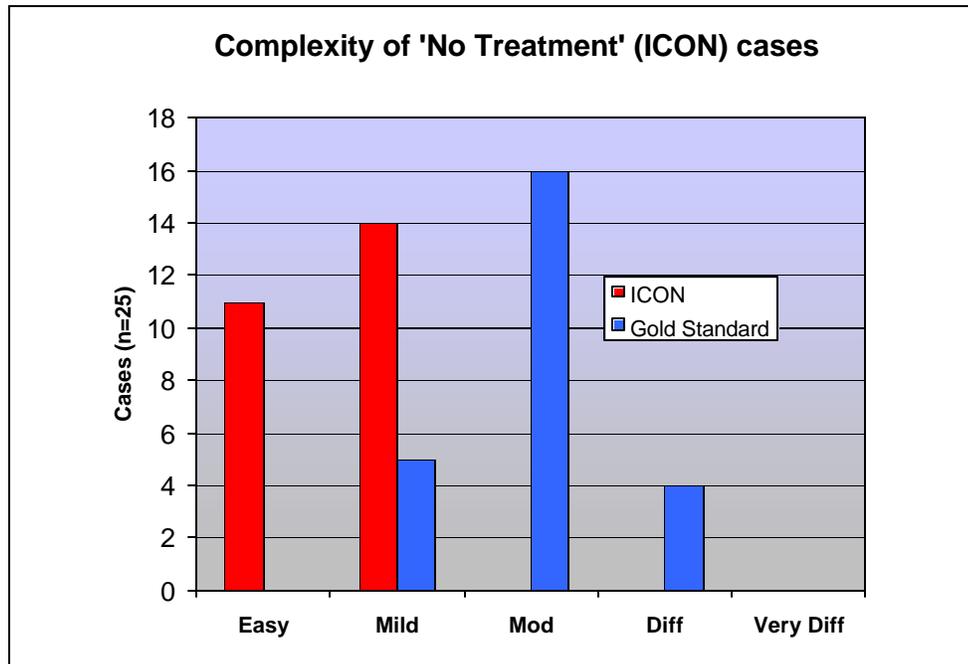


Figure 8. Complexity for ‘No treatment’ cases .

An appreciable underscoring, by the ICON, of *complexity* for the 25 cases scored is evident (figure 8).

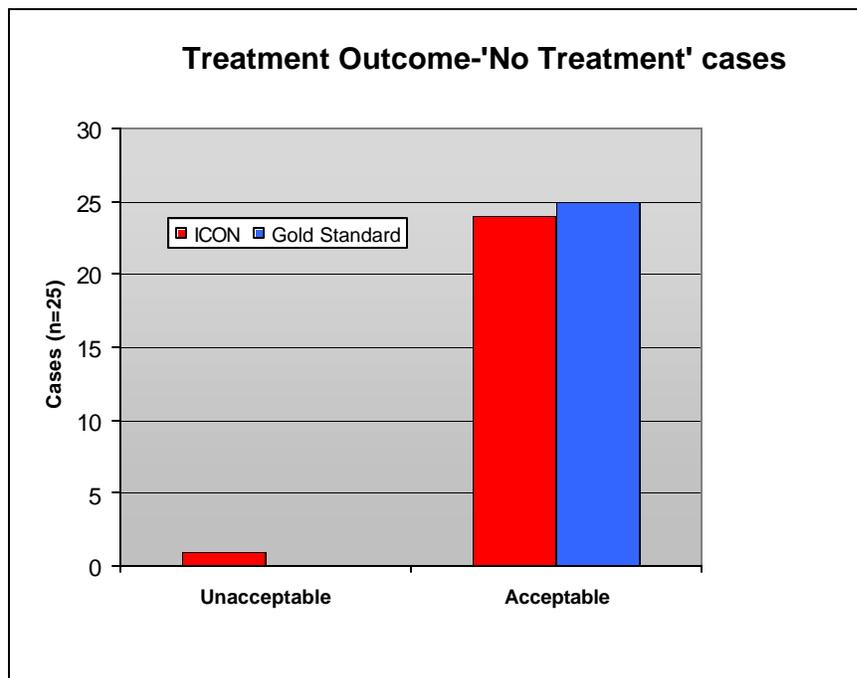


Figure 9. Treatment Outcome

The distribution of the scoring for *outcome* was similar to that found for all 100 cases (figure 9).



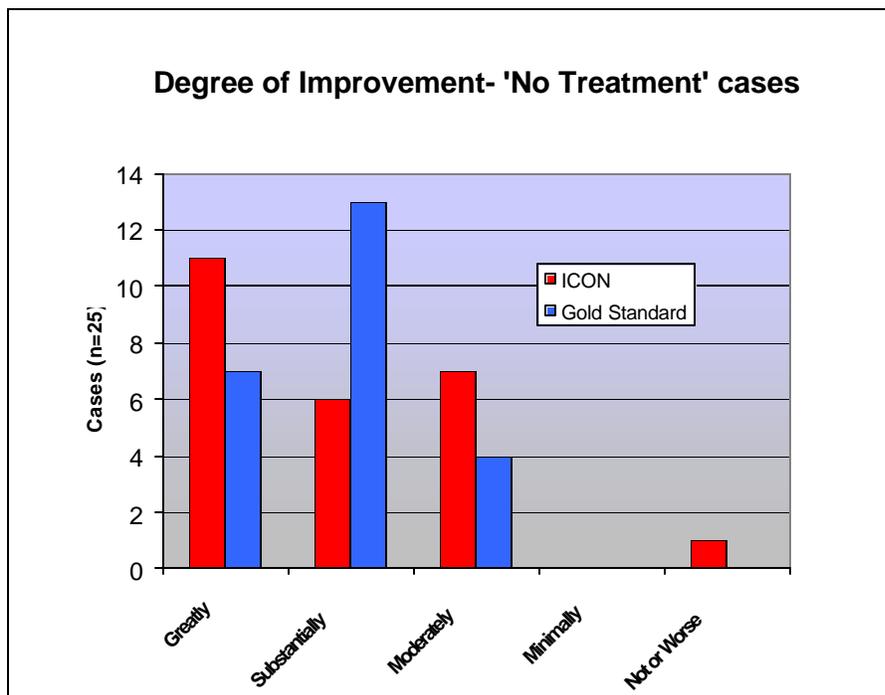


Figure 10. Degree of Improvement



The ICON scored highly for the categories 'greatly and substantially' improved (figure 10).

Objective Two

To assess ease of use and simplicity of the ICON

To determine the ease of use and the simplicity of the ICON, an additional 25 pre- and post-treatment study casts were examined. The ICON scores obtained from the researcher were then compared with those of 4 orthodontic specialists and 3 non-specialists (2 dentists and an oral hygienist), using the ICON for the first time.

This comparison revealed an agreement of 'fair to good' for all categories of the ICON except for *treatment outcome*, which showed excellent agreement of 100% (Table 8).

Table 8. Calibrated vs Specialist and Non-specialist

ICON	KAPPA
Treatment Need	0.59 **
Outcome	100% agreement
Complexity	0.53**
Degree of Improvement	0.59**

** fair to good agreement



Within the groups, specialists were compared to the non-specialists to test for level of agreement. Agreement for the ICON components, *treatment need* and *outcome* was high amongst all the examiners.

For *degree of improvement*, the kappa values showed a 'fair to good' agreement. Agreement for *Complexity* was poor (? = 0.35)(Table 9).

These results show a large variation between practitioners for the different components of the index.

Table 9. A comparison of Specialist vs Non-Specialist ICON scores

ICON	KAPPA
Treatment Need	0.75***
Outcome	100% agreement
Complexity	0.35*
Degree of Improvement	0.46**

* Poor agreement

** fair to good agreement

*** excellent agreement

The time taken for the practitioners to score the study models using the ICON was also noted. All the practitioners completed their measurements within 10 minutes of each other, the time ranging from 60 to 70 minutes.



Ease of use and simplicity of the index for orthodontic and non-orthodontic personnel is demonstrated by these results.

Objective Three

Applicability of the ICON scores obtained clinically and on study models

The objective was to assess the applicability of the ICON on patients and their corresponding study models. Sixty untreated patients from the orthodontic waiting list were initially identified, but only 52 study casts

could be obtained. These casts were then rated using the ICON and the results are presented below.

Treatment Need

Little variation for *treatment need* between the patient ICON and the cast ICON is observed (figure 11).

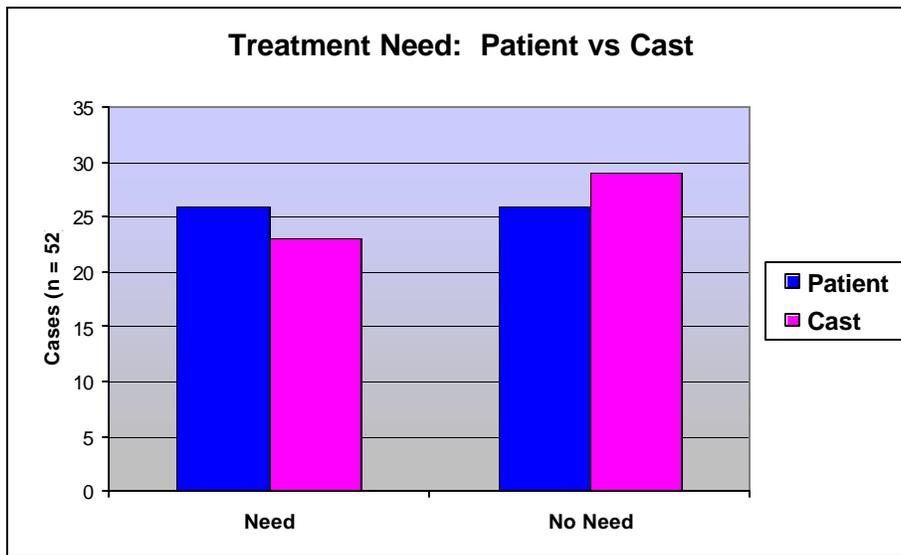


Figure 11. Treatment need. Patient vs cast

Complexity

The frequency distribution (Figure 12) of the scores obtained for the various categories was closely matched.

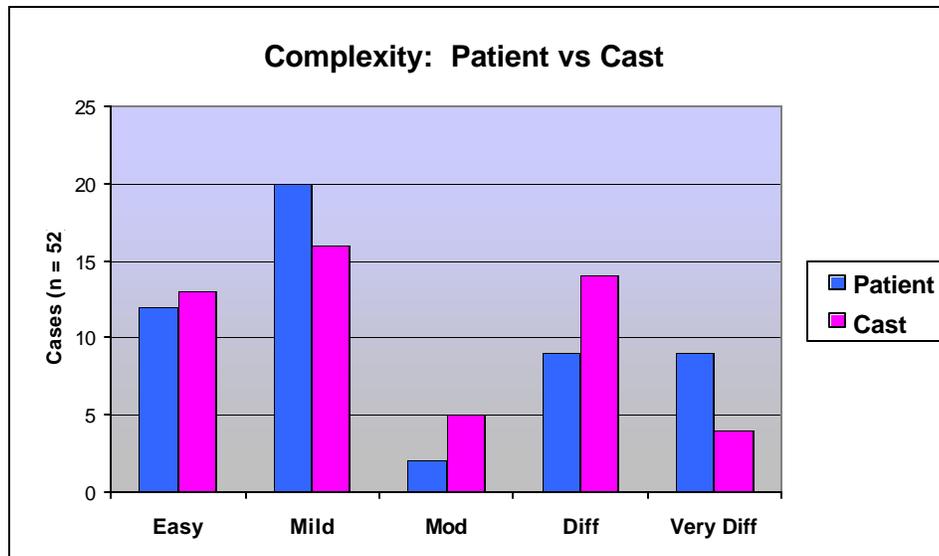


Figure 12. Complexity. Patient vs Cast

A high level of agreement was obtained for, crowding, crossbites, treatment need and complexity, whereas buccal relationships, aesthetics and overbite showed an agreement of 'fair to good' (Table 10).

Table 10. Patient ICON Score vs Cast ICON score

ICON	KAPPA
Aesthetics	0.72**
Crowding	0.79***
Crossbite	0.75***
Overbite	0.72**
Buccal relationship	0.61**
Treatment Need	0.88***
Complexity	0.84***

** fair to good agreement

*** excellent agreement

These results show that the ICON can be applied on the patient as well as the corresponding casts.



CHAPTER FIVE: DISCUSSION



The results will be discussed according to the 4 components of the ICON; treatment need, complexity, treatment outcome and degree of improvement.

Treatment Need

The IOTN scored 94% *treatment need* for the sample. The ICON scored 75% of the sample as needing treatment compared with the 100% of the gold standard. Firestone (2002) investigated the validity of the ICON as an index of orthodontic treatment. His values for sensitivity (94%), specificity (85%), positive predictive value (92%), negative predictive value (90%) and overall accuracy (91%) were appreciably different to those found in this study.



The aesthetic component of the IOTN (an integral feature of the ICON) has also been criticized for its poor ability to represent dentofacial imbalance in the antero-posterior plane, which is often associated with malocclusions (Evans and Shaw, 1987; Buchanan et al., 1994).

A factor contributing to the ICON underscoring these cases is that it does not score overjet, only overbite. This means that the index is blind to cases having well-aligned arches with a large overjet. Another factor is that the index scores Class I molar relationships the same as it does Class II and Class III relationships. Therefore a Class II division 1 malocclusion with well-aligned arches would not be scored as a 'treatment need' case.

Firestone (2002) showed that changing the cut-off point would directly influence the number of cases included in the *treatment need* category.

Reliability for *treatment need* was 'fair to good' ($\kappa = 0.63$), which can be attributed to the shortcomings of the aesthetic component of the scoring protocol.

Fox et al. (2002) found that the general performance of the ICON index in their study compared favourably with the IOTN in assessing *treatment need*.

In this study, the ICON lacked sensitivity when compared with the gold standard and thus its validity is questionable. However, the results of the intra-examiner values indicate that the ICON is a reliable index of *treatment need*.



Complexity

Complexity is a concept that is very difficult to define. A number of definitions have been proposed but have proven to be inadequate in one respect or another. Often something is classified as being complex that one would intuitively see as being simple, or alternatively an obviously complex phenomenon would not be labeled as such. Therefore what exactly defines a case as easy or difficult is not clear (Richmond et al., 2001).

There was a strong tendency for raters to score in the midmost category (moderately complex) and none scored in the first (easy) in assessing complexity as a measurement of treatment difficulty. It is assumed that this vagueness in definition therefore contributed to the raters being indecisive or unwilling to commit to a category. Another possible explanation is that treatment difficulty may mean different things to different clinicians for different cases.

Hence the statistical analysis of this component of the ICON yielded a 'poor' kappa value ($\kappa = 0.20$).

The weighted kappa values for intra-rater reliability was excellent ($\kappa = 0.80$), similar to the values reported by Savastano (2003).



Richmond et al. (1997), using the IOTN and PAR indices, studied the professional perception of orthodontic treatment complexity by sampling specialist and non-specialist practitioners in General Dental Services (GDS) in the United Kingdom (UK). They concluded that a number of confounding factors such as cost, number and length of appointments, age of patient at the start of treatment, and initial PAR score were all significantly associated with the judgment of difficulty. In addition, their study found no predictive factors prior to treatment to assess orthodontic treatment.

Complexity in this study was found to be reliable but achieved 'poor' validity.

The developers of the ICON have stated that further validation in this area is needed before the complexity assessment can be used to predict success (Daniels and Richmond, 2000).

Treatment Outcome

There was perfect agreement between the ICON and the gold standard. Therefore, *treatment outcome* was shown to have excellent validity.

The gold standard determined that all the cases needed treatment and that the outcome was acceptable. The ICON, scoring 25% of the cases as 'no treatment' (but nevertheless having been treated), determined that 97% of the cases had an acceptable outcome.

In the study by Savastano et al. (2003) agreement between the raters and the ICON scores was 'fair to good' ($\kappa = 0.50$). One possible explanation given for the lower-than-expected kappa value was the high prevalence of acceptable outcome scores as determined by the PAR index. Initial, not-final PAR scores were used to select the 100 subjects for this study. Therefore, the range of variability for treatment finishes was limited. An excellent level of agreement was shown for the rating of *treatment outcome* within intra-examiner reliability of 100%.

Treatment outcome was shown to be a reliable and valid component of the ICON.

Degree of Improvement

The agreement for the *degree of improvement* was “poor” ($\kappa = 0.34$) between ICON and the gold standard.

An excellent reliability ($\kappa = 0.84$) was found for this component of the index.

Savastano et al. (2003) reported low inter-rater reliability for *degree of improvement* and could not validate that component of the ICON. They concluded that a wider range of treatment finishes; including a greater number with unacceptable outcomes was needed in their sample.

As with *complexity*, *degree of improvement* is also poorly defined and raters appear to have different opinions as to what the different categories of improvement may signify.

An analysis of the “No Treatment” ICON group

Further analysis of the 25% cases identified by the ICON as not needing treatment was done. The index underscored *complexity* (figure 8), which may offer an explanation for it allocating a quarter of the cases to the “no treatment” category.

The distribution of the scoring was consistent to that found for the rest of the 75 cases and therefore did not reveal anything of significance (figure 9).

It is interesting to note that the ICON determined that a significantly high proportion of the cases were substantially or greatly improved. Notwithstanding the fact that it had identified these as not needing treatment. The question may be asked as to why cases not needing treatment could have become greatly improved?

Ease of use and Simplicity

The agreement between the researcher and the practitioners was 'fair to good' ($\kappa = 0.59$) whereas the specialist and non-specialist displayed high agreement ($\kappa = 0.75$). This discrepancy may be due to the researcher being calibrated and the relative inexperience of the practitioners in using the ICON.

However, these values indicate that the index is simple and easy to use.

Agreement between ICON scores for *complexity* between the researcher and the practitioners was 'fair to good' ($\kappa = 0.53$); between the specialist and non-orthodontic personnel agreement was 'poor' ($\kappa = 0.35$). This illustrates the difficulty in the definition of *complexity*.

Outcome between the groupings of the index was perfect (100%) proving this component of the ICON to be easy to use and simple.

Slightly better agreement was shown between examiner and practitioners ($\kappa = 0.59$) than within the practitioner group (0.46). This 'fair to good' agreement is further confirmation of the ease of use and simplicity of the ICON.

The time taken by the practitioners to score the models was within the range proposed by Daniels and Richmond (2000).

In summary, all the components ICON, excluding *complexity*, proved to be simple and easy to use.



Applicability to patient and study cast

Agreement for *treatment need*, and *complexity* between ICON scores obtained from patients and casts ($\kappa = 0.88$) was high, demonstrating that the index is applicable in determining orthodontic *treatment need* in both patients and study casts.

Weightings

Some indices, namely the HMAR, the OI and the DAI, make use of weightings to rank particular features in order of importance. These weightings use mathematical models based on previous experience. Weightings, however, suggest that certain features of malocclusion warrant more attention than others. There is little evidence to support this view (Jenny and Cons, 1998).

In the ICON, the aesthetic component is heavily weighted. This raises some concern, because it is the most criticized component (Evans and Shaw, 1987; McGuinness and Stephens, 1995; Trotman and Elsbach, 1996; Flores-Mir et al., 2004; Mattick et al., 2004). The shortcomings of the aesthetic component would then automatically be incorporated into the ICON and be compounded by the heavier weighting.

In this study the ICON was shown to be reliable. However the questions concerning the validity of the various components were raised. Despite these shortcomings the ICON has been touted as the index of choice.

The arbitrary cut-off points would benefit fund administrators as the under scoring of *treatment need* would exclude patients from treatment programmes.

Shortcomings of the ICON

In summary the following shortcomings were identified:

1. The terminology concerning the components of the index were inadequately explained.
2. The index uses the aesthetic component of the IOTN, which inherently has its own shortcomings.
3. The index ascribes too high a weighting to the aesthetic component.
4. Class I, II and III buccal relationships are given the same scoring.
5. Overjet is not measured,
6. The index disregards the lower arch.
7. Midlines are also not taken into consideration.

LIMITATIONS OF THE STUDY



All the cases rated had treatment done. It would be prudent to include cases not needing treatment to increase the range of occlusal characteristics that would be examined.

The concepts of Complexity, Degree of Improvement and Treatment outcome should have been explained to all the participants more adequately.

The different aspects of the study should have used a common sample.

CHAPTER SIX: CONCLUSION



CONCLUSION

In all respects, except for the components *complexity and degree of improvement*, the index performed well with respects to the ideal requirements of reliability, validity (of treatment outcome), ease of use and simplicity and applicability to patient and study casts.

RECOMMENDATIONS

1. The *treatment need* component needs to be reassessed in a follow-up study.
2. *Complexity and degree of improvement* should be more clearly defined.
3. It is also recommended that the weighting assigned to the aesthetic component be revised.



REFERENCES



A.A.O. Bulletin. St. Louis: American Association of Orthodontists; fall 1990.

Abdullah, M.S.B., Rock, W.P. (2001). Assessment of orthodontic treatment need in 5,112 Malaysian children using the IOTN and DAI indices. *Comm Dent Health*. 18, 242-248.

Ackerman, J.L., Proffit, W.R. (1969). The characteristics of malocclusion. A modern approach to classification and diagnosis. *Am J Orthod Dentofac Orthop*. 56: 443-454.

Albino, J. E., Lewis, E. A., Slakter, M. J. (1978). Examiner reliability for two methods of assessing malocclusion, *Angle Orthod*. 48:297-302.



Al Yami, E.A., Kuijpers-Jagtman, M., van 't Hof, M.A. (1998). Occlusal outcome of orthodontic treatment. *Angle Orthod*. 68: 439-444.

Andrews, L. (1972). The six keys to normal occlusion. *Am J Orthod Dentofac Orthop*. 6: 296-309.

Angle, E. H. (1899). Classification of malocclusion, *Dental Cosmos*. 41: 248-264, 350-357.

Ansai, T., et al. (1994). Prevalence of malocclusion in high school students in Japan according to the Dental Aesthetic Index. *Community Dent Oral Epidemiol.* 21: 303-305.

American Association of Orthodontists. (1993). Proceedings of the Orthodontic Indices Consensus Conference. St Louis. 114-117. (cited in Lindauer, 1998).

Baume, L.J., Horowitz, H.S., Summers, C.J., et al. (1973). A method for measuring occlusal traits developed by the FDI commission on classification and statistics for oral conditions (COCSTOC), *Int Dent J.* 23: 530-537.

Beglin, F.M., et al. (2001). A comparison of the reliability and validity of 3 occlusal indexes of orthodontic treatment need. *Am J Orthod Dentofac Orthop.* 120: 240-246.

Bezroukov, V., et al. (1979). Basic method for recording occlusal traits. *Bulletin of the world Health Organization.* 57(6): 955-961.

British Orthodontic Standards Working Party. (1986). Second report. *Br J Orthod.* 13: 165-173.

Brooke, P.H., Shaw, W.C. (1989). The development of an index of orthodontic treatment priority. *Eur J Orthod.* 11: 309-318.

Bjork, A., Krebs, A.A., Solow, B. (1964). A method for epidemiological registration of malocclusion. *Acta Odontol Scand.* 22: 27-41.

Buchanan, I.B., Downing, A., Stirrups, D.R. (1994). A comparison of the Index of Orthodontic Treatment Need applied clinically and to diagnostic records. *Br J Orthod.* 21 (2): 185-188.

Buchanan, I.B., Shaw, W.C., Richmond, S., O'Brien, K.D., Andrews, M. (1995). A comparison of the reliability and validity of the PAR Index and Summers' Occlusal Index. *Eur J Orthod.* 15: 27-31.

Burden, D.J. (1995). The ranking of dental aesthetics. *Br J Orthod.* 22: 259-61.



Carlos, J.P. (1970). Evaluation of indices of malocclusion. *Int Dent J.*; 20: 606-617.

Carlos, J.P., Ast, D.S. (1966). An evaluation of the HLD Index as a decision making tool. *Public Health Rep.* 81(71): 621-626.

Case, C.S. (1921). *Technics and principles of Dental Orthopedia.* Chicago, CS Case Co, 15-50. (cited in Ackerman and Proffit, 1969).

Cons, N.C., Jenny, J., Kohout, F.J. (1986). *DAI: The Dental Aesthetic Index,* College of Dentistry, University of Iowa, Iowa City.

Cons, N.C., Jenny, J., Kohout, F.J. (1989). Utility of the Dental Aesthetic Index in Industrialized and Developed Countries. *J of Public Health Dentistry*. 49 (3): 163-166.

Dawjee, S.M., Ackerman, A., Shaw, W.C. (2002). An aesthetic component of the IOTN for black subjects. *SADJ*. 57(7): 258-263.

Daniels, C., Richmond, S. (2000). The development of the index of complexity, outcome and need (ICON). *Br J Orthod*. 27: 149-62.

De Guzman, et al. (1995). The validation of the Peer Assessment Rating Index for malocclusion severity and treatment difficulty. *Am J Orthod Dentofac Orthop*. 107: 172-176.



De Mûelenaere, K.R., Coetzee, C.E., Ackerman, A. (1998). The treatment need of a group of senior dental students as assessed by the IOTN and PAR indices. *Jnl of the South African Dental Association*. 53: 185-191.

Draker, H.L. (1960). Handicapping labio-lingual deviations: A proposed index for public health purposes. *Am J Orthod* 46: 295-305.

Elderton, R.D., Clarke, J.D. (1984). Orthodontic treatment in the General Dental Services assessed by the occlusal index. *Brit J of Orthod*. 10: 178-186.

Elderton, R.D., Clarke, J.D. (1983). An investigation of treatment in the GDS for patients with Class II division 1 malocclusion. *Brit J of Orthod.* 11: 2-8.

Elderton, R.D., Clarke, J.D. (1992). Orthodontic treatment in the general dental service assessed by the occlusal index. *Brit Dent J.* 170: 107-112.

Elsasser, W.A. (1953). Studies of dentofacial morphology II. Orthometric analysis of facial pattern. *Am J Orthod.* 39: 193-200.

Evans, R., Shaw, W. (1987). Preliminary evaluation of an illustrated scale for rating dental attractiveness. *Eur J Orthod.* 9: 314-318.

Fleiss, J.L., Cohen, J. (1973). The equivalence of weighted kappa and the intra-class correlation coefficient as measures of reliability. *Educ Psychol Meas.* 33: 613-619.

Flores-Mir, C., Major, P.W., Salazar, F.R. (2004). Self-perceived orthodontic treatment need evaluated through 3 scales in a university population. *Journal of Orthodontics.* 31: 329-334.

Firestone, A.R., Beck, M., Berglin, F.M., Vig, K.W.L. (2002). Validity of the index of Complexity, Outcome and Need (ICON) in determining orthodontic treatment need. *Angle Orthod;* 72: 15-20.

Freer, E., Freer, T.J. (1999). Variations in treatment need using four screening methods. *Aust J Orthod.* 15: 214-218.

Fox, N.A. (1993). The first 100 cases: a personal audit of orthodontic treatment assessed by the PAR (Peer Assessment Rating) Index. *Brit Dental J.* 174: 290-297.

Fox, N.A., Daniels, C., Gilgrass, T. (2002). A comparison of the Index of Complexity Outcome and Need (ICON) with the Peer Assessment Rating (PAR) and the Index of Orthodontic Treatment Need (IOTN). *Brit Dental Journal.* 192: 225-230.

Ghafari, J., et al. (1989). Longitudinal evaluation of the Treatment Priority Index (TPI). *Am J Orthod Dentofac Orthop.* 96: 382-389.



Grainger, R.M. (1961). Malocclusion Severity Estimate, p9-11, Burlington Orthodontic Research Center, Progress report, Series VI.

Graveley, J. F., Johnson, O.B. (1974). Angle's classification of malocclusion: an assessment of reliability. *Br J Orthod.* 1: 79-86.

Gray, A.S., Demirjian, A. (1977). Indexing occlusions for dental public health programs. *Am J of Orthod.* 72: 191-197.

Grewe, J.M., Hagan, D.V. (1972). Malocclusion indices. A comparative evaluation, *Am J of Orthod.* 61: 286-294.

Hamden, A.M., Rock, W.P. (1999). An appraisal of the Peer Assessment Rating (PAR) Index and a suggested new weighting system. *Eur J Orthod.* 21: 181-192.

Han, H., Davidson, W. (2001). A useful insight into 2 occlusal indexes: HLD(Md) and HLD(CalMod). *Am J Orthod Dentofac Orthop.* 120: 247-253.

Hermanson, P.C., Grewe, J.M. (1970). Examiner variability of several malocclusion indices. *Angle Orthod* 40(3): 219-225.

Hlongwa, P., et al. (2004). Orthodontic treatment needs: Comparison of two indices. *SADJ* 59 (10): 421-424.



Ingervall, B., Ronnerman, A. (1975). Index for need of orthodontic treatment. *Odont Revy.* 26: 59-82. (cited in Järvinen, 2001).

Jago, J.D. (1974). The epidemiology of dental occlusion, a critical appraisal. *J Public Health Dent.* 34: 80-93.

Jamison, H.C., McMillan, R.S. (1966). An Index of Malocclusion for Use in Multiphasic Screening and Epidemiological Investigations. *Ala J Med Sci* 3: 154-158. (cited in Draker, 1960).

Järvinen, S., Väättäjä, P. (1987). Variability in assessment of need for orthodontic treatment when using certain treatment need indices. *Community Dent Oral Epidemiol.* 15: 245-248.

Jenny, J., Cons, N.C., Kohout, F.J., Jakobsen, J. (1993). Predicting handicapping malocclusion using the Dental Aesthetic Index (DAI). *Int Dent J.* 43: 128-131.

Jenny, J., Cons, N.C. (1996). Establishing malocclusion severity levels on the Dental Aesthetic Index (DAI) scale. *Aust Dental J.* 41(1): 43-46.

Keay, P.A., Freer, T.J., Basford, K.E. (1993). Orthodontic treatment need and the Dental Aesthetic Index. *Aust Orthod J.* October: 4-7.



Kerr, W.J.S., Buchanan, I.B., McColl, J.H. (1993). Use of the PAR Index in assessing the effectiveness of removable orthodontic appliances. *Brit J Orthod.* 20: 351-357.

Kinaan, B.K., Burke, P.H. (1981). Quantitative assessment of the occlusal features. *Br J Orthod.* 8: 149-156.

Lewis, E.A., Albino, J.E., Canut, J.J., Tedesco, L.A. (1993). Reliability and validity of clinical assessments of malocclusion. *Am J Orthod.* 81: 473-477.

Linder-Aronson, S. (1974). Orthodontics in the Swedish public dental health service. *Trans Eur Orthod Soc.* 233-240.

Lindauer, A.J., Thresher, A.A., Baird, B.W., Sheats, R.D., Rebellato, J. (1998). Orthodontic treatment priority: a comparison of two indices. *J of Clin Pediatr Dent.* 22: 125-131.

Lobb, W.K., Ismail, A.I., Andrews, C.L., Spracklin, T.E. (1994). Evaluation of orthodontic treatment using the Dental Aesthetic Index. *Am J Orthod Dentofac Orthop.* 106: 70-75.

Mattick, C.R., Gordon, P.H., Gillgrass, T. J. (2004). Smile aesthetics and malocclusion in UK teenage magazines assessed using the Index of Orthodontic Treatment. Need (IOTN) *Journal of Orthodontics.* 31: 17–19.



McGorray, S. (2002). Sensitivity, Specificity and Related Concepts. *Seminars in Orthodontics.* June 8 (2).

McGuinness, N.J., Stephens, C.D. (1995). An introduction to Indices of Malocclusion. *Dent Update.* 8: 17-23.

McKnight, M.M., Daniels, C.P., Johnson, L.E. (1998). A retrospective study of two-stage treatment outcomes assessed with two modified PAR indices. *Angle Orthod.* 68(6): 521-526.

Mclain, J.B, Proffit, W.R. (1995). Oral health status in the United States: prevalence of malocclusion. *Journal of Dental Education*. 49: 386-396.

Moyers, R.E. (1988). *Classification and Terminology of Malocclusion*.

O'Brien, K. *Handbook of Orthodontics*. Chapter 9, 4th Edition. Chicago. p183-195.

Otuyemi, O. D., Jones, S.P. (1995). Methods of Assessing and grading malocclusion: A Review. *Aust Orthod J*. 14: 21-27.

Otuyemi, O. D., Noar, J.H. (1995). Variability in recording and grading the need for orthodontic treatment, using the handicapping malocclusion assessment record, occlusal index and dental aesthetic index. *Community Dentistry and Oral Epidemiology*. 24: 222-224.



Pickering, E.A., Vig, P. (1976). The Occlusal Index to assess orthodontic treatment. *British Journal of Orthodontics*. 2: 47-51.

Richmond, S., et al. (1992). The development of the PAR Index (Peer Assessment Rating): reliability and validity. *Eur J Orthod*. 14: 125-139.

Richmond, S., et al. (1992). The PAR index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvement and standards. *Eur J Orthod*. 14: 180-187.

Richmond, S., O'Brien, K.D. (1996). Health gain in orthodontics: a comparison between the general dental services and hospital service. *Comm Dent Health*. 13: 128-132.

Richmond, S., et al. (2001). A 2Center Comparison of Orthodontist's Perceptions of Orthodontic Treatment Difficulty. *Angle Orthodontist*. 17(5): 404-410.

Salzmann, J.A. (1968). Handicapping malocclusion assessment to establish treatment priority. *Am J Orthod*. 54: 749-765.

Shaw, W.C., Richmond, S., O'Brien, K.D., Brook, P., Stephens, C.D. (1991a). Quality control in Orthodontics: Indices of Treatment need and treatment standards. *British Dental Journal*. Feb 9: 107-112.



Shaw, W.C., Roberts, C.T. (1993). The use of occlusal indices in assessing the provision of orthodontic treatment by the hospital orthodontic service of England and Wales. *Brit J Orthod*. 20: 25-35.

So, L.L.Y., Tang, E.L.K. (1993). A comparative study using the occlusal index and the index of orthodontic treatment need. *Angle Orthod*. 63: 57-64.

Summers, C.J. (1966). A method for identifying and scoring occlusal disorders: The Occlusal Index. Doctor of Public Health Dissertation, University of Michigan. (Cited in Summers, 1971).

Summers, C.J. (1971). The occlusal index: A system for identifying and scoring occlusal disorders. *Am J Orthod.* 59: 553-567.

Summers, C.J. (1972). Test for validity for indices of occlusion. *Am J Orthod.* 62 (4): 428-429.

Tang, E.L.K., Wei, S.H.Y. (1993). Recording and measuring malocclusion: A review of the literature. *Am J Orthod Dentofac Orthop.* 103: 344-351.

Tarvit, D.J., Freer, T.J. (1998). Assessing Malocclusion-the time factor. *Brit J Orthod.* 25 (1): 31 -34.



Trotman, A., Elsbach, G.H., (1996). Comparison of malocclusion in preschool black and white children. *Am J Orthod Dentofac Orthop.* 110: 69-72.

Turner, S. (1990). Occlusal Indices Revisited. *Brit J Orthod.* 17: 197-203.

Van Kirk, L.E., Pennell, E.H. (1959). Assessment of malocclusion in population groups. *Am J Orthod.* 45: 752-758.

World Health Organization (1966). An International methodology for epidemiological studies of oral diseases. Manual No. 5: Epidemiological studies of periodontal diseases, first draft, Geneva. (cited in Turner, 1990).

World Health Organization (1989). International Collaborative study of oral health outcomes (ICS II), document 2-oral data collection instrument and examination criteria. Geneva: WHO; 13-27.



APPENDIX

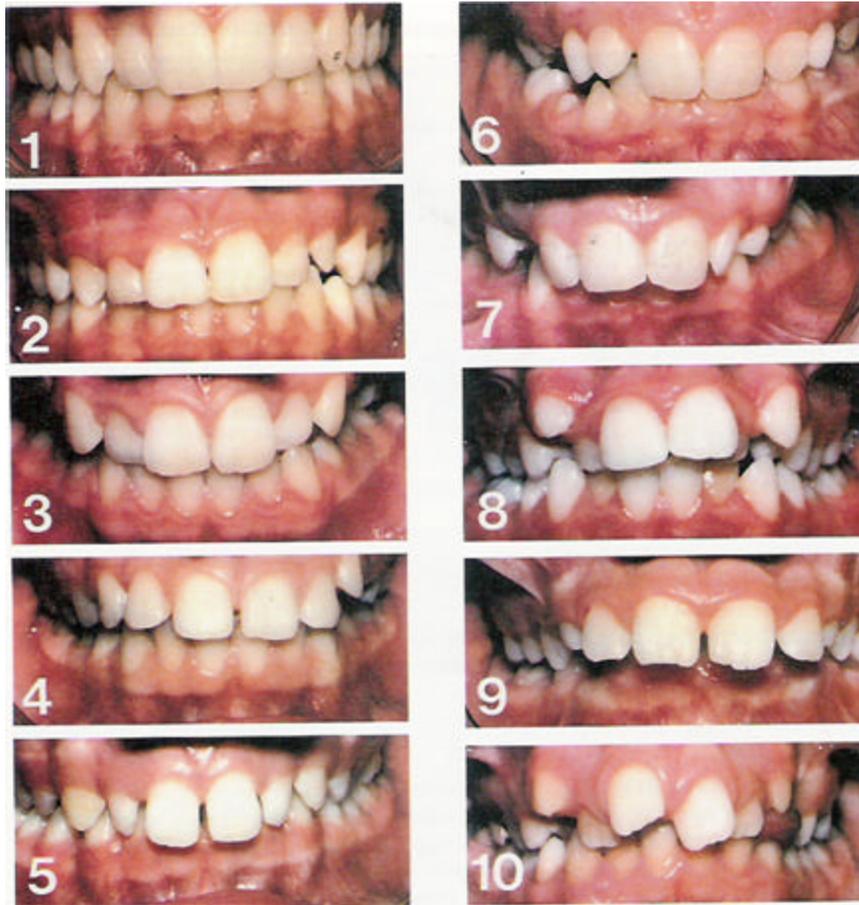


APPENDIX I

Table 11. ICON scoring protocol

	Score	0	1	2	3	4	5
Aesthetic	1-10 As judged using IOTN AC						
Upper arch Crowding	Score only the highest trait either spacing or crowding Transverse relationship of cusp to cusp or worse	Less than 2 mm	2.1 to 5 mm	5.1 to 9 mm	9.1 to 13 mm	13.1 to 17 mm	> 17 mm or impacted teeth
Upper spacing		Up to 2 mm	2.1-5 mm	5.1-9mm	>9mm		
Crossbite		No Crossbite	Crossbite				
Incisor open bite	Score only the highest trait either open bite or overbite Lower incisor coverage	Complete Bite	Less than 1mm	1.1-2mm	2.1-4mm	>4mm	
Incisor overbite		Up to 1/3 tooth	1/3-2/3 coverage	2/3 up to full covered	Fully covered		
Buccal segment anteroposterior or	Left and right added together	Cusp to embrasure relationship only, Class I, II, or III	Any cusp relation up to but not including cusp to cusp	Cusp to cusp relationship			

The Aesthetic Component of the IOTN



The index contains five components , all of which must be scored.

1. Dental Aesthetics

1. The dental aesthetic component of the IOTN (Shaw et al., 1991a) is used.

2. The dentition is compared to the illustrated scale and a global attractiveness match is obtained without attempting to closely match the malocclusion to a particular picture on the scale. The scale works best in the permanent dentition.

3. The scale is graded from 1 for (the most attractive to 10) the least attractive dental arrangement. Once this score is obtained it is multiplied by the weighting of 7.

2. Upper Arch Crowding/Spacing



This variable attempts to quantify the tooth to tissue discrepancy present in the upper arch or the presence of impacted teeth in both arches.

The sum of the mesio-distal crown diameters is compared to the available arch circumference, mesial to the last standing tooth on either side. This may require the use of a millimeter rule for accuracy, but with practice can be estimated by eye.

No estimation is made to account for the curve of Spee or the degree of incisor inclination. Once the crowding/spacing discrepancy has been worked out in mm, it is reduced on to the ordinal scale (0-5) using the categories shown in the Table 13.

Note that an impacted tooth in either the upper or lower arch, immediately scores the maximum for crowding. A tooth must be unerupted to be defined as impacted.

An unerupted tooth is defined as impacted under the following conditions:
if it is ectopically placed or impacted against an adjacent tooth (excluding third molars but including supernumerary teeth);
when less than 4 mm of space is available between the contact points of the adjacent permanent teeth.

Retained deciduous teeth (without a permanent successor) and erupted supernumerary teeth should be scored as space unless they are to be retained to obviate the need for a prosthesis. In the transitional stages average canine and premolar widths can be used to estimate the potential crowding. Suggested averages are 7 mm for the premolar and lower canine and 8 mm for upper canine respectively. The presence of erupted antimeric teeth allows more accurate estimation for this purpose. Spacing due to teeth lost to trauma and exodontia is also counted.

Post-treatment spaces created to allow prosthetic replacements should match the antimeric tooth width. Discrepancy between such spaces and the antimeric tooth can be counted as excess spacing or crowding, whichever is appropriate. The use of the index to assess spacing in relation to retained deciduous teeth demands that the fate of the deciduous teeth is known before the index can be applied.

Once the raw score has been obtained it is multiplied by the weighting 5.

2. Crossbite

A normal transverse relationship in the buccal segments is observed when the palatal cusps of the upper molar and premolar teeth occlude, preferably into the occlusal fossa of the opposing tooth, or at least between the lingual and buccal cusp tips of the opposing tooth. Crossbite is deemed to be present if a transverse relation of cusp-to-cusp or worse exists in the buccal segment. This includes buccal and lingual crossbites consisting of one or more teeth, with or without mandibular displacement.

In the anterior segment, a tooth in crossbite is defined as an upper incisor or canine in edge-to-edge or lingual occlusion.

Where a crossbite is present in the posterior or anterior segments or both, the raw score of 1 is given which is multiplied by the weighting of 5.

When there is no crossbite the score for this trait is 0.



4. Anterior Vertical Relationship

This trait includes both open bite (excluding developmental conditions) and deep bite. If both traits are present only the highest scoring raw score is counted. Positive overbite is measured at the deepest part of the overbite on incisor teeth.

Open bite may be measured with an ordinary mm rule to the mid incisal edge of the most deviated upper tooth.

The raw score obtained is multiplied by 4.

5. Buccal Segment Antero-posterior Relationship

The scoring zone includes the canine, premolar and molar teeth. The antero-posterior cuspal relationship is scored according to the protocol given in the Table 13 above for each side in turn. The raw scores for both sides are added together and then multiplied by the weighting 3.

Use of the index to assess treatment need.

To use the index to assess treatment need the pre-treatment study models are examined and occlusal traits are scored according to the protocol below. The five occlusal trait scores are then multiplied by their respective weightings and summed (Table 14). If the summary score is greater than 43, treatment is indicated.

Table 12. ICON index variables, weightings and cut-off values for treatment need and outcome decisions

Occlusal Trait	ICON index weightings
IOTN Aesthetic Component	7
Left + Right Buccal antero-posterior	3
Upper arch Crowding	5
Overbite	4
Crossbite	5
Treatment need cut-off	43
Treatment outcome cut-off	31

An example of a case being scored.

Case 1	Pretreatment score	Post-treatment score	Weighting
Aesthetic Component	5	2	7
Upper arch crowding	5	0	5
Crossbite	0	0	5
Overbite	0	1	4
Buccal antero-posterior relationship	4	1	3
Weighted score	72	21	
Treatment?	Yes	N/A	
Complexity	Difficult	N/A	
Improvement	N/A	Substantially improved	
Outcome	N/A	Acceptable	



Use of the index to assess treatment outcome

To assess treatment outcome, apply the index scoring method to the post-treatment models only. If the summary score is less than 31 the outcome is acceptable.

Use of the index to assess treatment complexity

To assess treatment complexity a, five-point scale is used via the cut off points for the 20 percentile intervals, using ranges given in Table 15 from the pre-treatment models.

Table 13. ICON index complexity cut-off values

Complexity grade	Score Range
Easy	Less than 29
Mild	29 to 50
Moderate	51 to 63
Difficult	64 to 77
Very Difficult	greater than 77

Use of the index to assess the degree of improvement

To assess the degree of improvement multiply the post-treatment score by 16, and subtract the result from the pre-treatment score. The ranges in Table 5 are used to assign a grade.

Table 14. Degree of improvement Scale

Pre-treatment – 4 (Post-treatment) ICON index score ranges, for ratings of treatment improvement	
Improvement grade	Score Range
Greatly improved	>-1
Substantially improved	-25 to –1
Moderately improved	-53 to –26
Minimally improved	-85 to –54
Not improved or worse	<-85

APPENDIX II

QUESTIONNAIRE FOR EXPERTS

Cast No.....

Please tick the appropriate box.

TREATMENT NEED

Y	N
---	---

OUTCOME

Acceptable	Not Acceptable
------------	----------------

COMPLEXITY

Easy	Mild	Moderate	Difficult	Very Difficult
------	------	----------	-----------	----------------

DEGREE OF IMPROVEMENT



Greatly Improved	Substantially Improved	Moderately Improved	Minimally Improved	Not Improved Or worse
------------------	------------------------	---------------------	--------------------	-----------------------

APPENDIX III
STATISTICAL ANALYSIS

Table 15. A 2-by-2- contingency Table of Decisions to treat or not to treat for ICON vs. Expert Opinion

EXPERT OPINION	ICON SCORE	
	No treatment (> 42)	Treatment need (< 4 2)
No treatment		
Treatment need		

From these comparisons, the following values will be calculated for the index: sensitivity, specificity, positive and negative predictive values, accuracy (percentage agreement), chi-squared and kappa statistic. Sensitivity is the percentage of all cases needing treatment that the index identified as needing treatment. Specificity is the percentage of all cases not needing treatment that the index identified as not needing treatment. Positive and negative predictive values are the percentage of cases that the index identified as needing (positive) or not needing (negative) treatment that in fact need or do not need treatment. Accuracy of the test is an overall summary of how well the test classifies those patients with and without the disease (Berglin et al., 2001).

The methods for calculating these measures and the positive and negative predictive values are given in Table 18.

Table 16. Calculation of validity

		Disease Status		
		Present	Absent	Total
Screening Test	Positive	a	b	a + b
	Negative	c	d	c + d
	Total	a + c	b + d	a + c + b + d

a = no. of true positive,

b = no. of false positives

c = no. of false negatives

d = no. of true negatives

- **Sensitivity** - probability of a positive test in people with the disease
= $a/(a + c)$
- **Specificity** - probability of a negative test in people without the disease = $d/(b + d)$ 
- **Positive predictive value** - probability of the person having the disease when the test is positive = $a/(a + b)$
- **Negative predictive value** - probability of the person not having the disease when the test is negative = $d/(c + d)$ (Beaglehole et al. 1993).
- **Accuracy** – is the overall summary of how well the test classifies those patients with or without the disease.

