

**Aus der Orthopädischen Universitätsklinik Tübingen  
Ärztlicher Direktor: Professor Dr. N. Wülker**

**Untersuchungen und Vergleich über die Zuverlässigkeit  
von Klassifikationen zur Einteilung der idiopathischen  
Skoliose**

**Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Medizin**

**der Medizinischen Fakultät  
der Eberhard-Karls-Universität  
zu Tübingen**

**vorgelegt von  
Dr. med. dent. Susanne Kluba, geb. Schubert  
aus Halle/ Saale**

**2005**

Dekan:	Professor Dr. C. D. Claussen
1. Berichterstatter:	Professor Dr. N. Wülker
2. Berichterstatter:	Professor Dr. K.-K. Dittel

Diese Arbeit ist meinen Eltern gewidmet.

## Abkürzungen

a.p.:	anterior-posterior
CSVL:	central sacral vertical line
GWS:	Wirbelsäulenganzaufnahme
DM:	double major (Lenke-Kurventyp 3)
DT:	double thoracic (Lenke-Kurventyp 2)
L:	Lumbal
Max.:	Maximum
Min.:	Minimum
mind.:	mindestens
MT:	main thoracic (Lenke-Kurventyp 1)
nom.:	nominal
SV:	stable vertebra
Th:	Thorakal
TL/L	thoracolumbar/lumbar (Lenke-Kurventyp 5)
TL/MT:	thoracolumbar/lumbar – main thoracic (Lenke-Kurventyp 6)
TM:	triple major (Lenke-Kurventyp 4)
u.a.:	unter anderem
U1:	Untersucher 1= Oberarzt
U2:	Untersucher 2= Assistenzarzt
U3:	Untersucher 3= Ärztin im Praktikum
VDS	ventrale Derotationsspondylodese
WK:	Wirbelkörper
z.B.:	zum Beispiel

# 1 Inhaltsverzeichnis

<b>Kapitel</b>		<b>Seite</b>
<b>1.</b>	<b>Einleitung</b>	7
<b>2.</b>	<b>Fragestellung</b>	8
<b>3.</b>	<b>Material und Methoden</b>	
3.1	Patientengruppen und Untersuchungsmaterial	9
3.2	Untersuchungsaufbau	10
3.3	Vorgehen bei den Messungen	11
3.4	Klassifikationen	
3.4.1	King-Klassifikation	14
3.4.2	Lenke-Klassifikation	15
3.4.2.1	Kurventyp	16
3.4.2.2	Lumbar spine modifier	17
3.4.2.3	Sagittal thoracic modifier	18
3.5	Statistische Auswertung	19
<b>4.</b>	<b>Ergebnisse</b>	
4.1	Allgemeine Ergebnisse	22
4.2	Intraobserver-Vergleich	
4.2.1	Auswertung Untersucher 1	24
4.2.2	Auswertung Untersucher 2	24
4.2.3	Auswertung Untersucher 3	25
4.2.4	Zeitaufwand und Intraobserver-Reliabilität im Untersuchervergleich	26
4.3	Interobserver-Vergleich	
4.3.1	Reliabilität bei Übereinstimmung aller drei Untersucher	28

4.3.2	Reliabilität der Teilschritte der Lenke-Klassifikation	29
4.3.3	Lernkurven	31
4.3.4	Interobserver-Reliabilität verschiedener Untersucherpaare	34
<b>5.</b>	<b>Diskussion</b>	
5.1	Allgemeines	37
5.2	King-Klassifikation	38
5.3	Lenke-Klassifikation	40
5.4	Einfluss der Cobb-Winkelmessung	41
5.5	Einfluss der Erfahrung unterschiedlicher Untersucher	43
5.6	Einfluss Kappa-Statistik	45
5.7	Zuverlässigkeit anderer orthopädischer Klassifikationen	47
5.8	Schlussfolgerungen und Ausblick	47
<b>6.</b>	<b>Zusammenfassung</b>	50
<b>7.</b>	<b>Literaturverzeichnis</b>	51

# 1 Einleitung

Klassifikationen sind geeignete Mittel, Informationen zu ordnen. Im Gegensatz zu Nomenklaturen sind sie gröbere Einteilungen, die zugunsten der Übersichtlichkeit einen gewissen Informationsverlust in Kauf nehmen. Es wird Ähnliches, das jeweils festgelegten Kriterien entspricht, einer Klasse zugeordnet (Krämer 1999, Krämer und Maichl 1993).

Klassifikationen sind für die Diagnostik und Therapie von Krankheiten sowie für die wissenschaftliche Arbeit bedeutend. Heterogenität von Begriffsdefinitionen erschwert die wissenschaftliche Vergleichbarkeit und die Entwicklung von Standards in Diagnostik und Therapie. Außerdem sollten Klassifikationen die Arbeit im klinischen Alltag erleichtern, indem sie unabhängig vom Ausbildungsstand des Arztes eine Einteilung erlauben. Gerade in der Orthopädie und Traumatologie finden sich eine Vielzahl klinischer Klassifikationen. Eine wichtige und weit verbreitete orthopädische Klassifikation ist z.B. die Weber-Klassifikation (Weber 1972).

Das Wissen eines Fachgebietes unterliegt ständigen Änderungen durch den Fortschritt medizinischer Forschung. Deshalb ist neben einer übersichtlichen Struktur, die möglichst einfache und schnelle Anpassung an neue Erkenntnisse eine wichtige Anforderung an medizinische Klassifikationen.

Lenke et al. haben mit der Entwicklung einer neuen Skolioseklassifikation versucht, dieser Anforderung gerecht zu werden (Lenke et al. 2001). Die vorliegende Arbeit beschäftigt sich mit dem Vergleich dieser Klassifikation mit der älteren, etablierten Einteilung der Skoliosen nach King (King et al. 1983). Beide sind der von der Scoliosis Research Society entwickelten Klassifikation angelehnt. Die Einteilung erfolgt anhand von Röntgenbildern und basiert auf der Lokalisation und Form der Krümmungen als Unterscheidungsmerkmal.

## 2. Fragestellung

Ziel der vorliegenden Untersuchung war die Überprüfung von Zuverlässigkeit und Anwendbarkeit von Klassifikationen zur Beurteilung der idiopathischen Skoliose. Dabei wurde die King-Klassifikation (King et al. 1983) mit einer neueren von Lenke (Lenke et al. 2001) vorgeschlagenen Einteilung verglichen. Im Rahmen der Arbeit sollten insbesondere folgende Fragestellungen näher untersucht werden:

1. Reproduzierbarkeit beider Klassifikationen (Intraobserver-Zuverlässigkeit)  
Inwieweit kommen die Untersucher bei zeitlich unabhängiger doppelter Beurteilung der Fälle zur selben Typzuordnung? Gibt es Unterschiede der untersucherspezifischen Fehleranfälligkeit zwischen der King- und der Lenke-Klassifikation? Sind diese abhängig vom Ausbildungsstand bzw. der Erfahrung des jeweiligen Untersuchers?
2. Reliabilität (Interobserver-Zuverlässigkeit) der King- und Lenke-Klassifikation  
Wie hoch ist die Zuverlässigkeit der beiden Verfahren bezüglich der Übereinstimmung aller Untersucher bei der Typzuordnung? Sind spezielle Teilschritte der Lenke-Klassifikation anfälliger für Fehler als andere? Beeinflusst der Ausbildungsstand die Quote der Übereinstimmungen? Ist ein Lerneffekt im Verlauf der Untersuchungen eingetreten?
3. Klinische Bedeutung Lenke-Klassifikation  
Inwieweit ist die Lenke-Klassifikation komfortabel in den klinischen Alltag zu integrieren? Welche Vor- und Nachteile besitzt die Lenke-Klassifikation gegenüber der King-Klassifikation?



### **3. Material und Methoden**

#### **3.1 Patientengruppen und Untersuchungsmaterial**

Eingeschlossen wurden 60 Patienten mit idiopathischen Skoliosen. Sekundäre Skoliosen wurden ausgeschlossen. Alter und Geschlechtszuordnung unterlagen bei der Auswahl der Patienten keiner Restriktion. Entsprechend der Häufigkeit des Auftretens der idiopathischen Skoliose stellen jüngere (13-18 Jahren) und weibliche Patienten die Mehrheit in der Untersuchungsgruppe dar. Alle Fälle stammen aus dem Patientengut der Orthopädischen Universitätsklinik Tübingen und waren dort aufgrund der Skoliose operativ behandelt worden.

Untersucht wurden präoperativ angefertigte Röntgenbilder. Wirbelsäulenganzaufnahmen (GWS) in vier verschiedenen Projektionen bildeten einen Untersuchungssatz (Abbildung 1). Die a.p.-Aufnahme und das seitliche Röntgenbild wurden stehend angefertigt. Die Bending-Funktions-Aufnahmen wurden im Liegen durchgeführt. Der Patient beugt dabei den Rumpf in der Frontalebene jeweils maximal seitlich nach rechts und links. Nach Korrektur in den Funktionsaufnahmen ist eine Beurteilung der Rigidität der Haupt- und Nebenkrümmungen möglich (Lenke et al., 2001; siehe auch Kapitel 3.4.2.1, Seite 16)

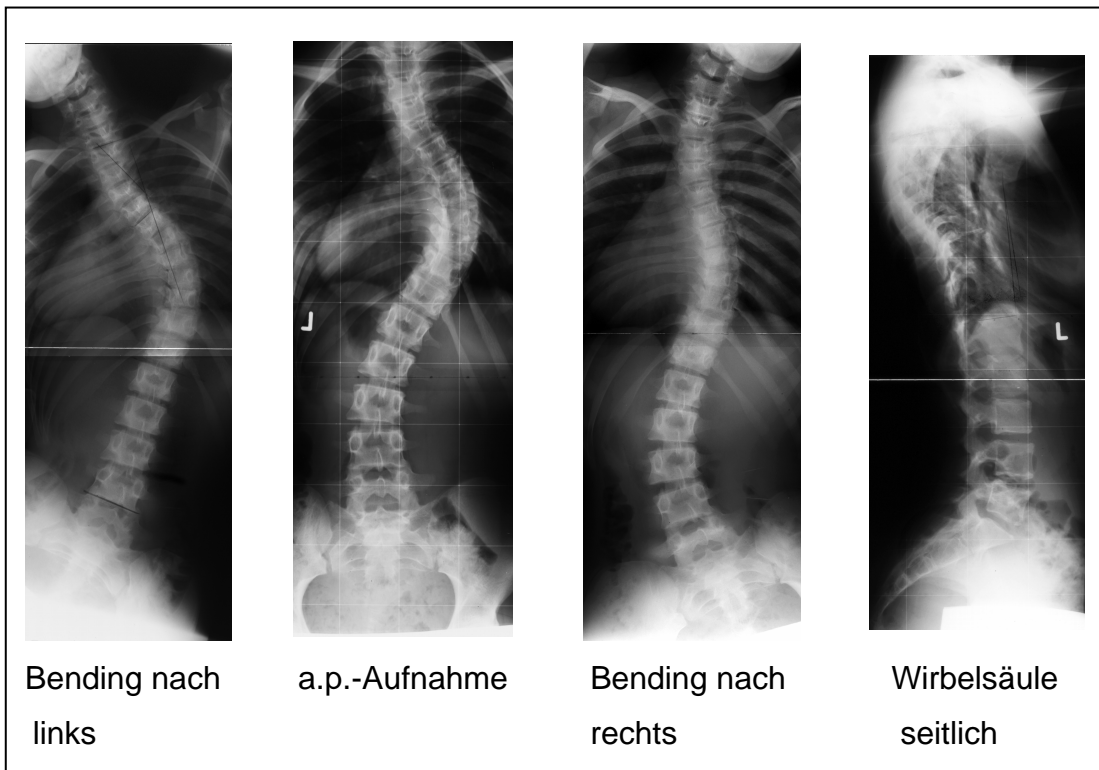


Abbildung 1: Projektionen untersuchter Röntgenaufnahmen

### 3.2 Untersuchungsaufbau

In der vorliegenden Studie wurden Röntgenbilder von Patienten mit idiopathischen Skoliosen nach dem Schema von King et al. (King et al., 1983) und entsprechend der Lenke-Einteilung (Lenke et al., 2001) klassifiziert. Dafür wurde die gesamte Untersuchungsgruppe von 60 Patienten in vier Blöcke mit jeweils 15 Patienten geteilt. Die Zuteilung zu einem Teilblock erfolgte zufällig. Die Beurteilung und Vermessung der Unterlagen fand für jeden Untersuchungsblock separat an unterschiedlichen Tagen statt.

Drei Ärzte unterschiedlichen Ausbildungsstandes werteten die Röntgenbilder zeitlich parallel und räumlich getrennt aus und ordneten sie entsprechend oben genannten Einteilungen zu. Bisher wurden beide getesteten Klassifikationen von keinem der Untersucher regelmäßig angewandt. Ein Untersucher verfügte über langjährige Erfahrung in der operativen Skoliogetherapie. (Untersucher 1). Des Weiteren nahm ein Arzt kurz vor Abschluss der Weiterbildung zum

Facharzt für Orthopädie teil (Untersucher 2). Er hatte bereits operative Erfahrung in der Wirbelsäulenchirurgie inklusive der Therapie von Skoliosen. Als dritter Untersucher war eine Ärztin mit eineinhalbjähriger klinischer Erfahrung in der Orthopädie in die Untersuchung eingebunden (Untersucher 3). Sie verfügte über keine operative Erfahrung in der Behandlung von Skoliosen.

Die Untersuchung wurde als Blindstudie durchgeführt. Die Anonymisierung der Patientendaten erfolgte durch Nummerierung der Röntgenbilder sowie Abkleben der Angaben auf den Röntgenbildern mit undurchsichtigem Heftpflaster. Die Zuordnung der Patientennummern und Röntgenbilder zu den Patienten war den Untersuchern unbekannt. Nach Abschluss eines jeden Auswertungsblockes und Entfernung der Abklebung wurden die Röntgenbilder wieder zu den Originalunterlagen der Patienten zurückgeführt. Eventuell vorhandene Markierungen oder Winkelangaben auf den Röntgenaufnahmen waren im Vorfeld der Untersuchung entfernt worden. Während der Auswertung vorgenommene Einzeichnungen wurden direkt im Anschluss von den Untersuchern wieder entfernt. Jeder Untersucher dokumentierte den Zeitaufwand für die Auswertung eines Patienten.

Im Anschluss an jeden Auswertungsblock erfolgte die Dokumentation und tabellarische Zusammenfassung der Ergebnisse. Danach trafen sich alle Untersucher zu einer Nachschau, in der die Röntgenbilder gemeinsam besprochen wurden. Bei Patienten mit uneinheitlichem Ergebnis in der Typzuordnung erfolgte eine erneute, diesmal gemeinsame Klassifizierung der Skoliose.

### **3.3 Durchführung der Untersuchungen**

Jedem Untersucher wurde ein eigenes Zimmer mit großem Röntgenschirm zugeteilt. Winkelmesser, weißer Röntgenstift, Kugelschreiber und eine Stoppuhr standen als Arbeitsmaterialien zur Verfügung. Des Weiteren bekamen alle Untersucher die Artikel von King et al. (1983) und Lenke et al. (2001) mit der Beschreibung der jeweiligen Klassifikation in Kopie ausgehändigt.

Bei jeder Zusammenkunft wurde ein Teilblock mit insgesamt 15 Patienten beurteilt. Es fanden jeweils acht Untersuchungsrunden in einer Sitzung statt. In den einzelnen Untersuchungseinheiten wurden pro Untersucher Unterlagen von 5 Patienten vorgestellt, die entweder nach King oder nach Lenke klassifiziert werden sollten. Die Ergebnisse von sechs dieser Einheiten dienten der Auswertung im Interobserver-Vergleich (Übereinstimmungen zwischen den Untersuchern). In den restlichen zwei Runden wurden den Untersuchern Röntgenbilder 5 weiterer Patienten zu einer zweiten Beurteilung (einmal nach King und einmal gemäß Lenke) vorgelegt. Durch dieses Verfahren hatte nach Abschluss aller vier Untersuchungsblöcke jeder der drei Untersucher 20 Patienten in beiden Klassifikationen doppelt ausgewertet. Die Daten der Doppelmessungen stellten die Grundlage in der Beurteilung der untersucherspezifischen Genauigkeit (Intraobserver-Vergleich) dar.

Sowohl die Verteilung der Unterlagen über einen speziellen Studienplan als auch die Verschlüsselung der Patientenangaben waren den Untersuchern nicht bekannt. Sie erhielten lediglich das Ergebnisformular mit Angabe der in dieser Runde durchzuführenden Klassifikation (Abbildung 2 und 3) sowie die anonymisierten Röntgenbilder. Die entsprechend Plan pro Patient zugeteilte Untersuchungsnummer von 1 bis 5 war für den Untersucher kenntlich gemacht. Anhand dieser Nummer konnte der Untersucher sein Ergebnis für diesen Patienten zusammen mit der ermittelten Beurteilungszeit in das Formular eintragen. Der Zeitaufwand pro Patient wurde definiert als der Zeitintervall ab Ende des Aufhängens aller vier Röntgenbilder auf dem Röntgenschirm bis zur endgültigen Zuordnung zu einer Kategorie durch den Untersucher.

**Untersucher:**

<b>Nr.</b>	<b>Zeit</b>	<b>King-Klassifikation</b>
1		
2		
3		
4		
5		

Abbildung 2: Untersucherformular für die King-Klassifikation

**Untersucher:**

<b>Nr.</b>	<b>Zeit</b>	<b>Lenke-Klassifikation</b>		
		Kurventyp (1-6)	lumbar spine modifier (A/B/C)	sagittal thoracic modifier (-/N/+)
1				
2				
3				
4				
5				

Abbildung 3: Untersucherformular für die Lenke-Klassifikation

Nach Abschluss einer Untersuchungseinheit wurden die Bilder und das ausgefüllte Formular dem Studienleiter übergeben, der unter Ausschluss der Untersucher die Neuverteilung für die nächste Runde vornahm.

### 3.4 Untersuchte Klassifikationen

#### 3.4.1 King-Klassifikation

Die radiografische Einteilung von King et al., 1983 für idiopathische Skoliosen unterscheidet 5 verschiedene Skoliosetypen. Die Beurteilung der Ausprägung der Skoliose basiert auf der Ermittlung der Cobb-Winkel (Cobb, 1948). Des Weiteren fließt in die Bestimmung des Subtypes dieser Klassifikation die Flexibilität der Krümmungen ein. Anhand der Bendingaufnahmen wird der Flexibilitätsindex bestimmt. Dieser Index ist definiert als die Differenz zwischen prozentualer Korrektur der lumbalen Krümmung abzüglich des Prozentsatzes einer erzielten Korrektur im thorakalen Bereich. Folgende Skoliosetypen sind durch King et al. beschrieben (Abbildung 4):

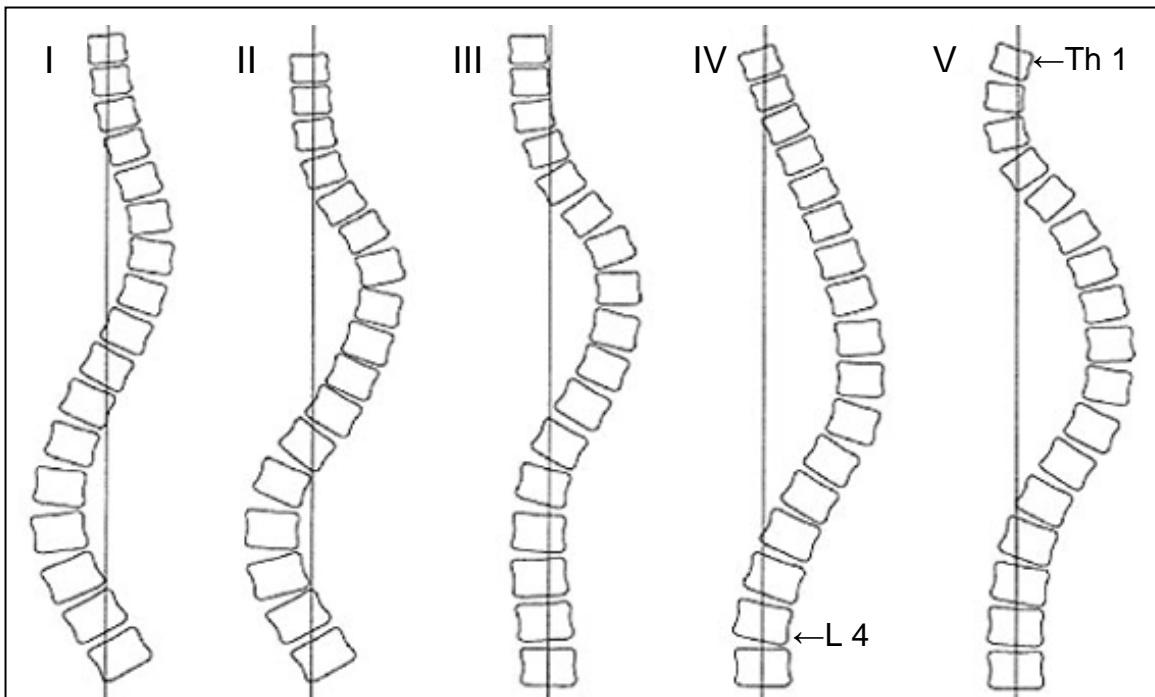


Abbildung 4: Subtypen der King-Klassifikation (mod. aus [www.medscape.com](http://www.medscape.com))

- Typ I: S-förmig. Thorakale als auch die lumbale Kurve überschreiten die Mittellinie, wobei die lumbale größer als die thorakale ist. Die lumbale Krümmung ist rigider als die thorakale (negativer Flexibilitätsindex in den Bendingaufnahmen)
- Typ II: S-förmig. Beide Krümmungen sind mittellinienüberschreitend. Die thorakale Krümmung ist größer und weniger flexibel (Flexibilitätsindex  $\geq 0$ ).
- Typ III: Thorakale Kurve. Die lumbale Krümmung überschreitet nicht die Mittellinie.
- Typ IV: Langbogige thorakale Krümmung. L 5 über dem Sakrum zentriert, während L 4 Richtung Krümmung geneigt ist.
- Typ V: Thorakale Doppelkrümmung. Th 1 in Konvexität der oberen, strukturellen Krümmung geneigt.

### **3.4.2 Lenke-Klassifikation**

Lenke et al. publizierten 2001 eine Einteilung idiopathischer Skoliosen anhand von Röntgenaufnahmen. Die Klassifikation unterscheidet insgesamt 42 Subtypen. Die Zuordnung zu einem Typ verläuft in 3 Teilschritten. Es werden Kurventypen (1-6) festgelegt, Veränderungen im Lumbalbereich beurteilt (lumbar spine modifier – A, B oder C) und das seitliche Profil beschrieben (sagittal thoracic modifier – -, N, +). Der Lenke-Typ wird in einer dreistelligen Zeichenkombination angegeben, z.B. 2A+. In Ermangelung entsprechender deutscher Begriffe werden die beiden letztgenannten Teilschritte in dieser Arbeit mit den englischen Bezeichnungen verwandt.

### 3.4.2.1 Kurventyp

Der Kurventyp wird anhand der Lokalisation (Abbildung 5), der Flexibilität und des Ausmaßes der Krümmung(en) festgelegt.

<u>Lokalisation:</u>	<u>Krümmungsscheitel:</u>
hochthorakal	WK Th 2 – WK Th 6
thorakal	WK Th 6 – Bandscheibe Th 11/12
thorakolumbal	WK Th12 – WK L1
lumbal	Bandscheibe L 1/2 – WK L 4

Abbildung 5: Krümmungslokalisation

Die Flexibilität der Kurven wird entweder durch die Größe der im Bending residuell vorhandenen Krümmung oder durch die Größe der Kyphose beschrieben. Eine Kurve gilt demnach als strukturell, wenn im Bending Cobb-Winkel  $\geq 25^\circ$  und /oder eine Kyphose von  $\geq 20^\circ$  nach Cobb gemessen werden. Außerhalb dieser Kriterien gelten alle Kurven als nicht strukturell. Dem Ausmaß der Krümmung wird durch die Unterscheidung von Haupt- und Nebenkrümmungen Rechnung getragen. Beide Formen können sowohl strukturellen als auch nicht strukturellen Charakter haben.

Insgesamt lassen sich anhand dieser Charakteristika sechs Kurventypen beschreiben:

1 (main thoracic – MT): thorakale, strukturelle Hauptkrümmung, andere Krümmungen nicht strukturell.

2 (double thoracic – DT): thorakale Hauptkrümmung und hochthorakale Nebenkrümmung strukturell, andere vorhandene Krümmungen nicht strukturell



- 3 (double major – DM): strukturelle thorakale und thorakolumbale/lumbale Kurve, thorakale Krümmung  $\geq$  thorakolumbaler/lumbaler Kurve, hochthorakale Krümmung nicht strukturell
- 4 (triple major – TM): alle Krümmungen strukturell, thorakale Hauptkrümmung
- 5 (thoracolumbar/lumbar – TL/L): thorakolumbale oder lumbale strukturelle Hauptkrümmung. hochthorakale/thorakale nicht strukturelle Nebenkrümmungen
- 6 (thoracolumbar/lumbar – main thoracic – TL/MT) strukturelle thorakolumbale oder lumbale Hauptkrümmung, strukturelle thorakale Nebenkrümmung (mind. 5°kleinerer Cobbwinkel)

#### 3.4.2.2 Lumbar spine modifier

Der lumbar spine modifier beschreibt die skoliotischen Veränderungen im lumbalen Teil der Wirbelsäule. Zur Ermittlung einer der drei möglichen Formen (A, B oder C) wird in der a.p.-Aufnahme eine Senkrechte zur Horizontalen mittig vom Sakrum nach kranial gezogen. Ein Beckenschiefstand  $> 2\text{cm}$  muss vor dem Röntgen ausgeglichen werden. Der Wirbelkörper, der durch diese Linie (central sacral vertical line = CSVL) am ehesten gleichmäßig geteilt wird, wird als stabiler Wirbelkörper (stable vertebra = SV) angesehen. Trifft dies für eine Bandscheibe zu, wird der kaudal gelegene Wirbelkörper als stable vertebra angenommen. Die Kriterien für die Zuordnung zu einer Form sind in Abbildung 6 dargestellt.

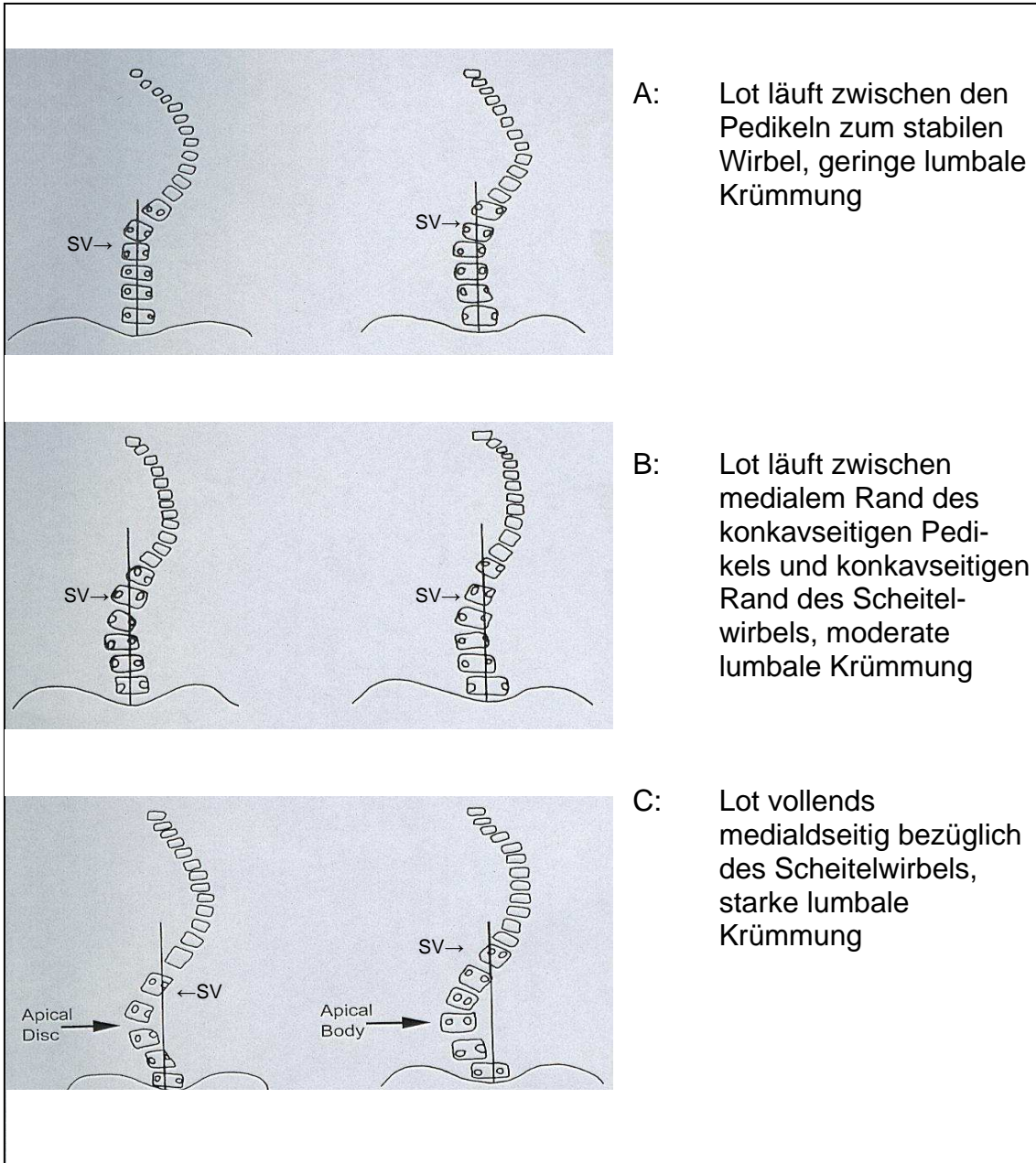


Abbildung 6: Ermittlung des lumbar spine modifier

### 3.4.2.3 Sagittal thoracic modifier

Im letzten Teilschritt der Lenke-Klassifikation wird das thorakale sagittale Profil beurteilt und wie folgt bezeichnet:

- : Kyphose Th 5-12 < +10° nach Cobb
- N: Kyphose Th 5-12 zwischen +10° und +40° nach Cobb
- +: Kyphose Th 5-12 > +40° nach Cobb

In Abbildung 7 ist beispielhaft die Ermittlung eines Subtypes entsprechend der Lenke-Klassifikation dargestellt.

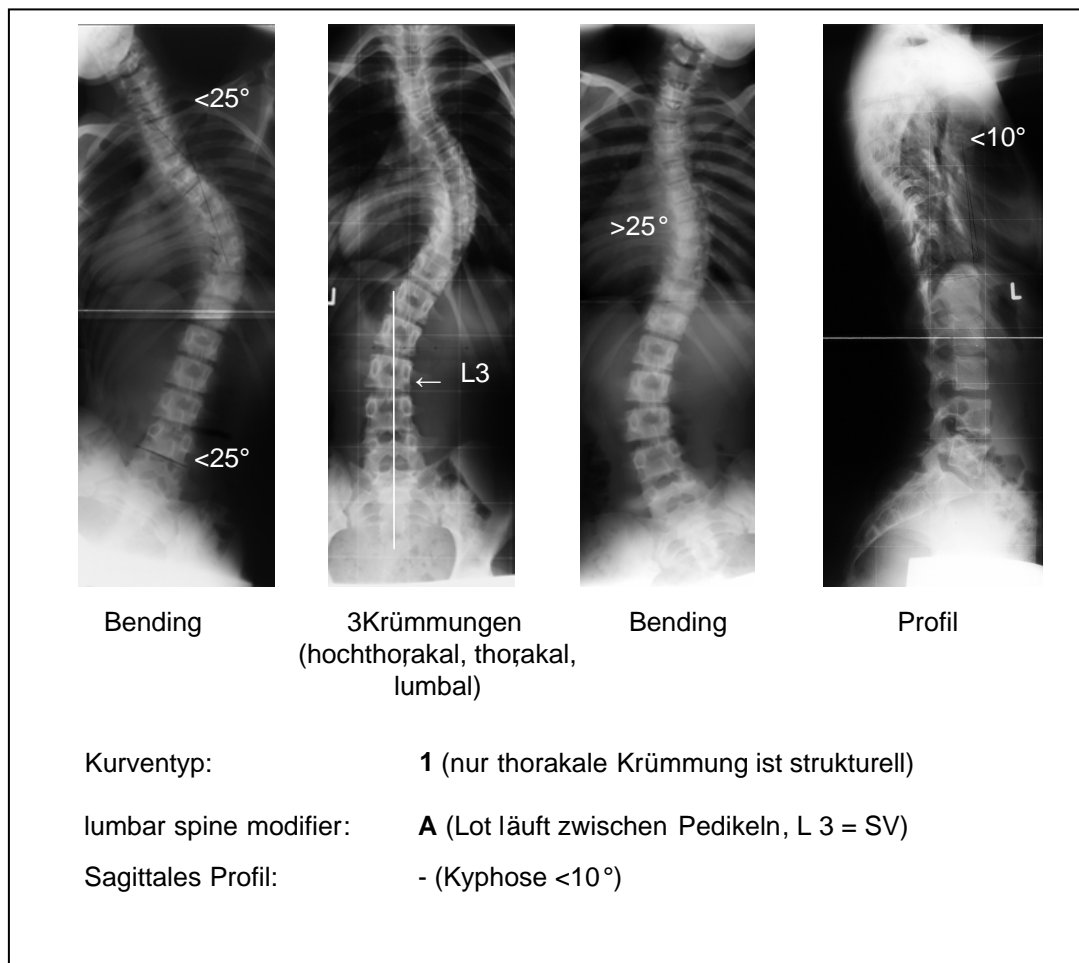


Abbildung 7: Lenke-Typ 1A-

### 3.5 Statistische Auswertung

Die Auswertung der Untersuchung erfolgte durch die nominale und prozentuale Angabe von Übereinstimmungen. Ein Teil beobachteter Übereinstimmung zwischen verschiedenen Untersuchern kann zufallsbedingt sein. Diesem

Umstand wurde durch Anwendung der Multi-Rater-Kappa-Statistik Rechnung getragen. Mit der Berechnung des Kappa-Wertes ( $\kappa$ ) wird das ermittelte Ergebnis um die Wahrscheinlichkeit zufälliger Übereinstimmungen korrigiert. Der Kappa-Koeffizient errechnet sich wie im Folgenden vereinfacht angegeben.  $P$  stellt die beobachteten Übereinstimmungen dar.  $P(\text{zufällig})$  wird über ein spezielles Rechenmodell ermittelt.

$$\kappa = \frac{P(\text{beob.}) - P(\text{zufällig})}{1 - P(\text{zufällig})}$$

Der  $\kappa$ -Koeffizient kann Werte zwischen 1 und -1 annehmen. Ein  $\kappa$  von 1 bedeutet volle Übereinstimmung, während bei  $\kappa=0$  die gefundenen Übereinstimmungen rein durch Zufall erklärbar sind. Negative Werte treten auf, wenn weniger Übereinstimmungen beobachtet wurden als durch Zufall zu erwarten wären.

Bei der Beurteilung der Präzision/ Reproduzierbarkeit medizinischer Klassifikationen anhand der Kappa-Statistik wurden die Bewertungsmaßstäbe von Landis und Koch (1977) und von Svanholm et al. (1989) angelegt (Abbildung 8)

Zuverlässigkeit	Svanholm et al., 1989	Landis and Koch, 1977
gering	<0,50	0-0,20
mäßig	0,50-0,75	0,21-0,40
akzeptabel		0,41-0,60
gut	>0,75	0,61-0,80
perfekt		0,81-1,00

Abbildung 8: Bewertung einer Klassifikation mittels  $\kappa$

Der Zeitaufwand für die Ermittlung der Klassifikationsergebnisse wurde mit dem Student-t-Test für unverbundene Stichproben mit ungleichen Varianzen geprüft. Dabei wurde eine Irrtumswahrscheinlichkeit von 1% ( $\alpha=0,01$ ) angenommen. Die Nullhypothese lautete jeweils: Die Messreihen unterscheiden sich nicht. Ein Unterschied wurde als signifikant betrachtet bei p-Werten  $<0,01$  und als hochsignifikant bei  $p < 0,001$ .

## **4. Ergebnisse**

### **4.1 Allgemeine Ergebnisse**

In der Gesamtheit der ausgewerteten Röntgenbilder nach King kamen alle 5 möglichen Subtypen vor. In der Bewertung entsprechend der Lenke-Einteilung waren dagegen von 42 möglichen Skoliosetypen nur 36 von den Untersuchern gewählt worden. Berücksichtigt man die in der gemeinsamen Nachschau der Blöcke vorgenommenen Korrekturen, reduziert sich die Anzahl der vorkommenden Lenke-Typen auf 23.

Der am häufigsten vorkommende Skoliosetyp war entsprechend der King-Klassifikation Typ 2 und bei Lenke der Subtyp 2AN. In den Lenke-Teilschritten waren zahlenmäßig der Kurventyp 1 und 2 mit 20 Patienten am stärksten vertreten. Der lumbar spine modifier C und der sagittal thoracic modifier N kamen häufiger vor als die jeweils anderen möglichen Ausprägungen.

Bei 4 von 60 Patienten war die Form der Skoliose durch die King-Einteilung nicht eindeutig zu beschreiben, so dass die Untersucher eine Klassifikation nach King für nicht möglich hielten. Ein Beispiel ist die in Abbildung 9 dargestellte thorakolumbale Skoliose. In der Klassifikation von Lenke konnte nur eine Skoliose nicht zugeordnet werden, bei der alle Krümmungen nach Lenke-Kriterien nicht strukturell waren (Abbildung 10).

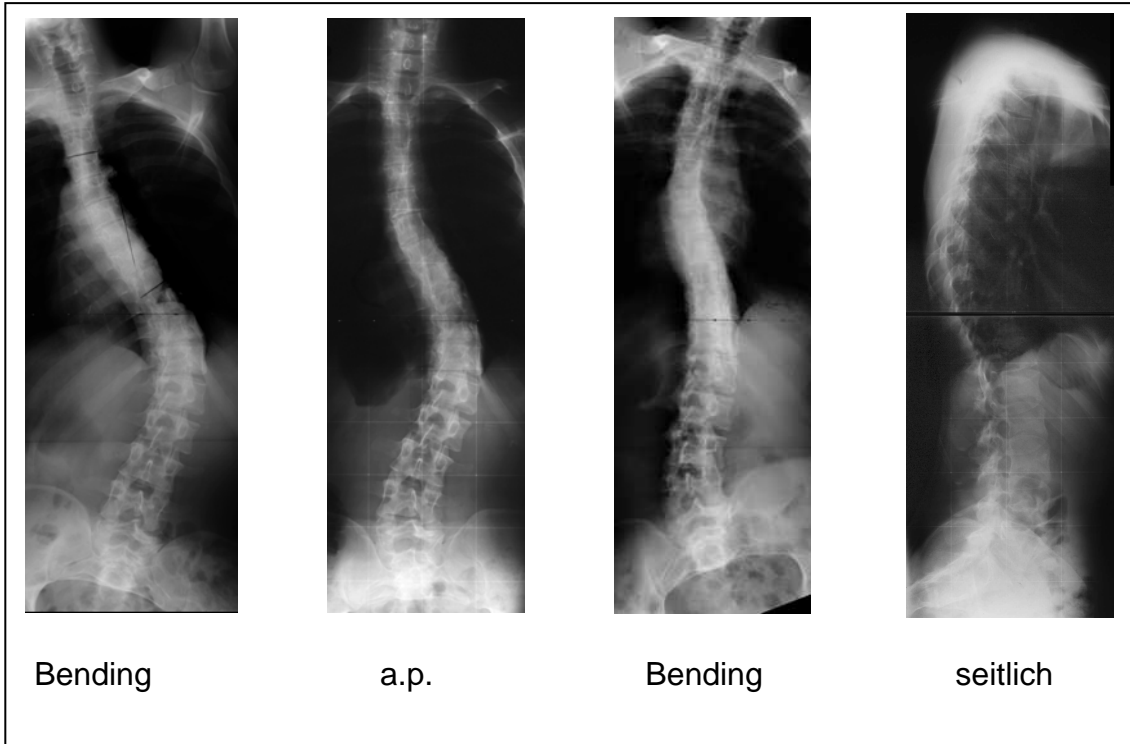


Abbildung 9: nach King nicht klassifizierbare thorakolumbale Skoliose

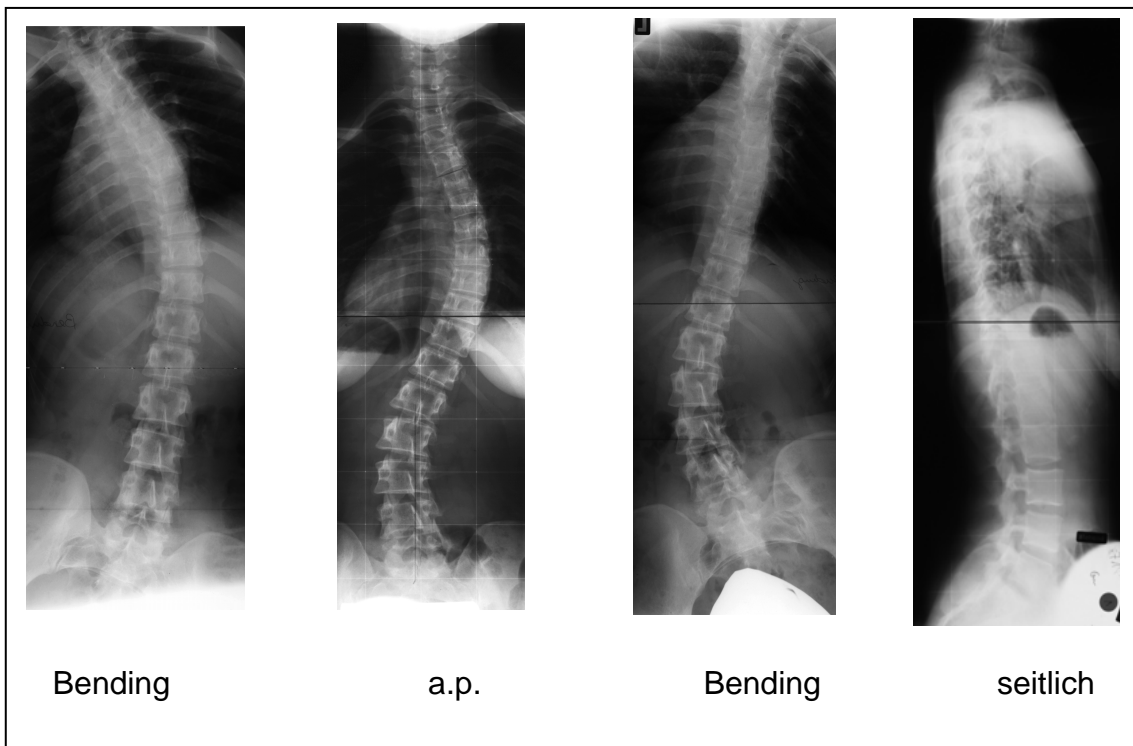


Abbildung 10: nach Lenke nicht klassifizierbare Skoliose

## 4.2 Untersucherspezifische Reproduzierbarkeit (Intraobserver-Vergleich)

### 4.2.1 Auswertung Untersucher 1

Die doppelte Beurteilung von 20 Fällen erbrachte für den Untersucher 1 die in Abbildung 11 dargestellten Ergebnisse.

Klassifikation	Übereinstimmung		Fehlerquote prozentual (%)	Kappa
	nominal	prozentual (%)		
King	<b>15</b>	<b>75</b>	25	<b>0,64</b>
Lenke	<b>11</b>	<b>55</b>	45	<b>0,51</b>
1. Kurventyp	16	80	20	0,69
2. lumbar spine modifier	19	95	5	0,91
3. sagittal thoracic modifier	15	75	25	0,44

Abbildung 11: Intraobserver-Vergleich Untersucher 1

Bei der Festlegung des endgültigen Klassifikationstypes bestanden bei der King-Klassifikation sowohl nominal als auch beim Kappa-Koeffizienten höhere Werte. Die Kappa-Werte der einzelnen Teilschritte der Lenke-Klassifikation lagen alle über dem entsprechenden Wert der Klassifikation nach King. Am höchsten war dieser Wert bei der Beurteilung des lumbar spine modifier.

### 4.2.2 Auswertung Untersucher 2

Klassifikation	Übereinstimmung		Fehlerquote prozentual (%)	Kappa
	nominal	prozentual (%)		
King	<b>16</b>	<b>80</b>	20	<b>0,74</b>
Lenke	<b>14</b>	<b>70</b>	30	<b>0,57</b>
1. Kurventyp	16	80	20	0,64
2. lumbar spine modifier	18	90	10	0,85
3. sagittal thoracic modifier	17	85	15	0,75

Abbildung 12: Intraobserver-Vergleich Untersucher 2



Der Untersucher 2 erzielte in den Doppelmessungen Übereinstimmungen zwischen 70 und 90 Prozent, was einer Variationsbreite des Kappa-Wertes von 0,57 bis maximal 0,85 entsprach (Abbildung 12). Die Reproduzierbarkeit lag bei der King-Klassifikation höher als bei Lenke. Bei der Lenke-Klassifikation konnte in 70% ein übereinstimmendes Ergebnis von erster und zweiter Beurteilung erzielt werden ( $\kappa=0,57$ ). Die Bestimmung des Kurventyps erzielte den niedrigsten Kappa-Koeffizienten mit 0.64. Alle Kappa-Werte für die Teilschritte waren größer als für den endgültigen Lenke-Typ mit  $\kappa=0,57$ .

#### 4.2.3 Auswertung Untersucher 3

Abbildung 13 zeigt die Ergebnisse des Intraobserver-Vergleichs für den Untersucher 3. Die errechneten Kappa-Werte befanden sich in Größenordnungen zwischen 0.31 und 0.70 und entsprachen damit einer Variationsbreite der Fehlerquote zwischen 65 und 15%. Die Beurteilung des sagittalen Profils im Rahmen der Lenke-Klassifikation erreichte den höchsten Kappa-Koeffizienten ( $\kappa=0,70$ ). Die Intraobserver-Zuverlässigkeit beider Klassifikationen war am geringsten von allen Untersuchern. Bei Festlegung der Klassifikationstypen wurden bei der Einteilung nach King doppelt so viele Übereinstimmungen erreicht wie bei Lenke. Dies entsprach einer Differenz der Kappa-Koeffizienten von 0.22.

Klassifikation	Übereinstimmung		Fehlerquote prozentual (%)	Kappa
	nominal	prozentual (%)		
King	<b>14</b>	<b>70</b>	30	<b>0,53</b>
Lenke	<b>7</b>	<b>35</b>	65	<b>0,31</b>
1. Kurventyp	12	60	40	0,42
2. lumbar spine modifier	14	70	30	0,51
3. sagittal thoracic modifier	17	85	15	0,70

Abbildung 13: Intraobserver-Vergleich Untersucher 3

#### 4.2.4 Zeitaufwand und Intraobserver-Reliabilität im Untersuchervergleich

Der zeitliche Aufwand für die Auswertung eines Patienten unterschied sich zwischen den Untersuchern und auch zwischen den beiden Klassifikationen. In Abbildung 14 sind die durchschnittlichen Auswertungszeiten pro Fall und Untersucher für beide Skolioseklassifikationen dargestellt. Für die King-Klassifikation benötigte der Untersucher 2 durchschnittlich mehr als dreimal so lang und der Untersucher 3 fast fünfmal so lang wie der Untersucher 1. Entsprechend ergab die statistische Testung zwischen Untersucher 1 und Untersucher 2 ( $p=1,16E-07$ ) und zwischen Untersucher 1 und Untersucher 3 ( $p=7,12E-05$ ) signifikante Unterschiede. Für die Lenke-Klassifikation erbrachte der t-Test im Zeitvergleich zwischen den Untersuchern hochsignifikante Unterschiede (p-Werte von  $1,40E-10$  bis  $8,96E-20$ )

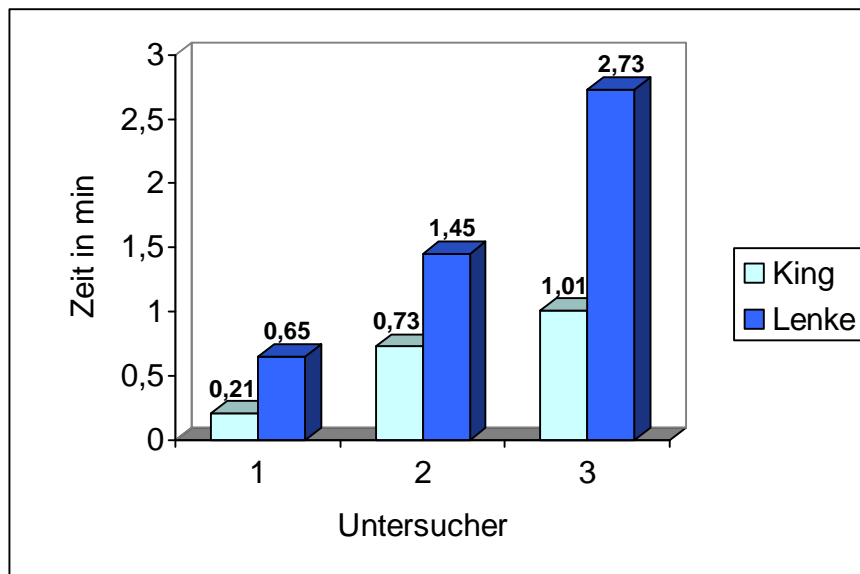


Abbildung 14: durchschnittliche Auswertungszeiten pro Patient

Die Einzelwerte der Zeitmessung zeigten in beiden Klassifikationen eine große Streuung. Die Variationsbreite fiel für den Untersucher 3 am höchsten aus. Der minimale Zeitaufwand betrug für die King-Klassifikation 3 Sekunden und reichte bis zu einem Maximalwert von 7,48 Minuten (Abbildung 15). Im Vergleich zwischen beiden Klassifikationen zeigten sich für alle Untersucher

hochsignifikante Unterschiede ( $p(U1)=1,40E-06$ ;  $p(U2)=9,44E-09$ ;  $p(U3)=1,24E-10$ ).

	King-Klassifikation		Lenke-Klassifikation	
	Min. in min	Max. in min	Min. in min	Max. in min
U1	0,02	1,36	0,10	3,01
U2	0,10	2,55	0,32	3,46
U3	0,03	7,48	0,29	6,20

Abbildung 15: Variationsbreite des Zeitaufwandes für die Auswertung

Die Intraobserver-Zuverlässigkeit zwischen den verschiedenen Untersuchern ist in Abbildung 16 dargestellt. Die Intraobserver-Zuverlässigkeit des Untersuchers 3 wies geringere Werte auf als die der beiden anderen Untersucher.

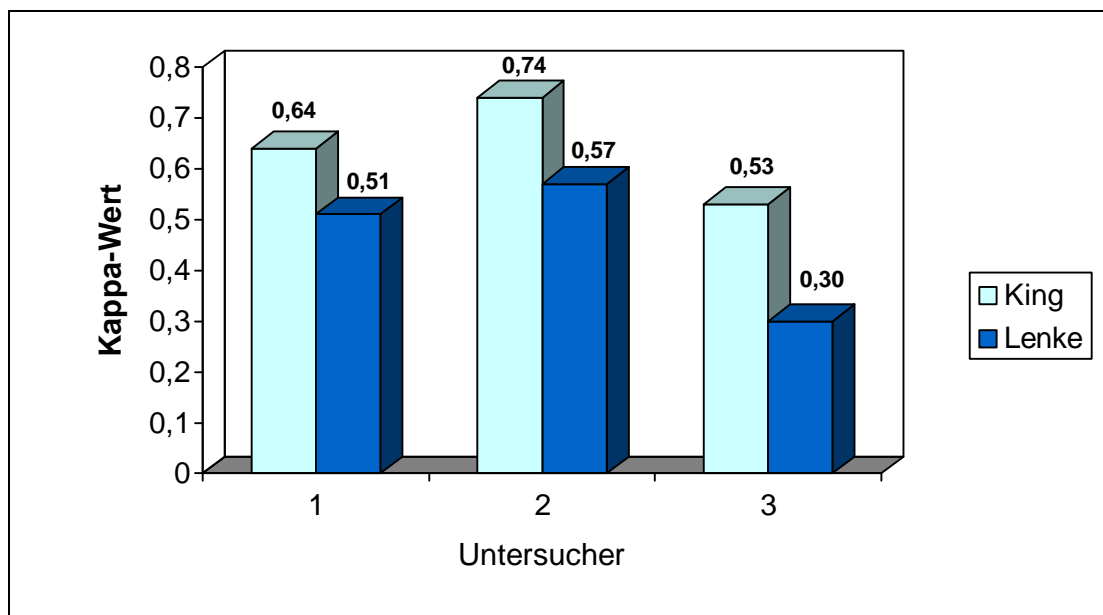


Abbildung 16: Untersuchervergleich Intraobserver-Reliabilität

### 4.3 Interobserver-Vergleich

#### 4.3.1 Reliabilität bei Übereinstimmung aller drei Untersucher

Im Interobserver-Vergleich wurden zunächst die Übereinstimmungen im Klassifikationsergebnis zwischen allen Untersuchern ermittelt. Abbildung 17 führt diese Ergebnisse vergleichend für beide Klassifikationen sowie aufgeschlüsselt nach Auswertungsblöcken auf. Bezogen auf alle vier Untersuchungsblöcke kamen in weniger als einem Viertel alle Untersucher zum selben Lenke-Klassifikationsergebnis. Die Rate an Übereinstimmungen war für die King-Klassifikation mehr als doppelt so hoch. Sie entsprach 51% aller untersuchten Patienten. Bei beiden Klassifikationen wurden die meisten Übereinstimmungen im letzten Untersuchungsblock erzielt.

Block	King		Lenke	
	nom.	%	nom.	%
alle	31	51,7	14	23,3
1	2	13,3	2	13,3
2	8	53,3	4	26,7
3	8	53,3	2	13,3
4	13	86,7	6	40,0

Abbildung 17: Übereinstimmungen aller Untersucher im Klassifikationsergebnis

Nach Berechnung des Kappa-Koeffizienten zeigten sich Veränderungen insbesondere für die King-Klassifikation (Abbildung 18). Während prozentual in den Untersuchungsblöcken 2 und 3 jeweils 53,3% Übereinstimmungen erzielt

wurden, lag der Kappa-Wert bei 0,43 für Block 2 und 0,51 für Block 3. Weitere Ausführungen dazu finden sich in der Diskussion (siehe auch Diskussion Seite 45).

Für die gesamte Untersuchung wurde für die Lenke-Klassifikation ein geringerer Kappa-Koeffizient erreicht als für die Klassifikation nach King. In der letzten Auswertungsrunde (Block 4) ergab sich für die King-Klassifikation die höchste Reliabilität ( $\kappa=0,79$ , Abbildung 18).

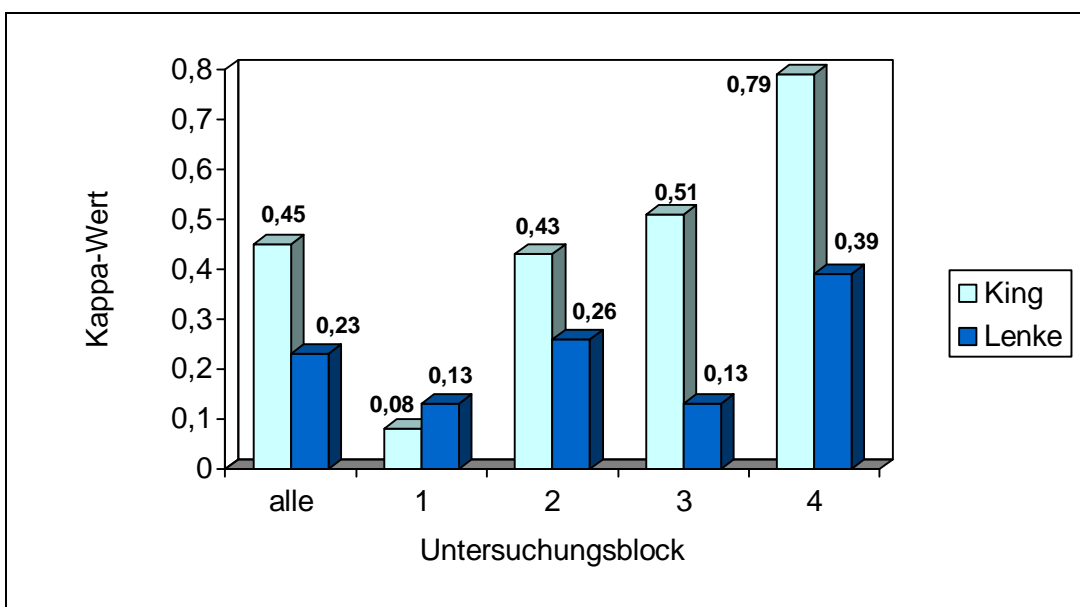


Abbildung 18: Interobserver-Vergleich (Kappa-Statistik)

#### 4.3.2 Reliabilität der Teilschritte der Lenke-Klassifikation

Die Skolioseklassifikation nach Lenke setzt sich aus drei Teilschritten zusammen. Zur genaueren Beurteilung dieser Klassifikation wurde nicht nur die Zuverlässigkeit der Gesamtklassifikation geprüft. Auch die Einzelschritte wurden separat ausgewertet (Abbildung 19).

Der lumbar spine modifier war mit einem Kappa-Wert von 0,51 (58,3% Übereinstimmungen aller Untersucher) weniger fehleranfällig als die beiden

anderen Teilschritte. Die Untersucher kamen bei der Bestimmung des sagittal thoracic modifier häufiger zum selben Ergebnis (56,7%) als bei der Festlegung des Kurventyps (53,3%). Der Kappa-Koeffizient des Kurventyps ist höher ( $\kappa=0,50$ ) als der des sagittal thoracic modifier ( $\kappa=0,45$ ).

Block	Kurventyp			lumbar spine modifier			sagittal thoracic modifier		
	$\kappa$	nom.	%	$\kappa$	nom.	%	$\kappa$	nom.	%
alle	0,50	32	53,3	0,51	35	58,3	0,45	34	56,7
1	0,36	6	40,0	0,42	7	46,7	0,36	6	40,0
2	0,61	10	66,7	0,59	10	66,7	0,35	8	53,3
3	0,33	6	40,0	0,41	7	46,7	0,58	11	73,3
4	0,63	10	66,7	0,65	11	73,3	0,46	9	60,0

Abbildung 19: statistische Auswertung der Lenke-Teilschritte

Im Verlauf der Untersuchung zeigten sich für alle Einzelschritte der Lenke-Klassifikation schwankende Werte im Vergleich der einzelnen Untersuchungseinheiten (Abbildung 19).

Alle drei Teilschritte wiesen separat betrachtet eine höhere Zuverlässigkeit aus, als die Gesamtklassifikation. Der Kappa-Wert lag zwischen 0,22 und 0,28 höher als für die Gesamtklassifikation. Das entspricht einer Erhöhung des  $\kappa$  von 195,7% bis zu 221,7% in den Teilschritten. Abbildung 20 verdeutlicht diesen Zusammenhang grafisch.

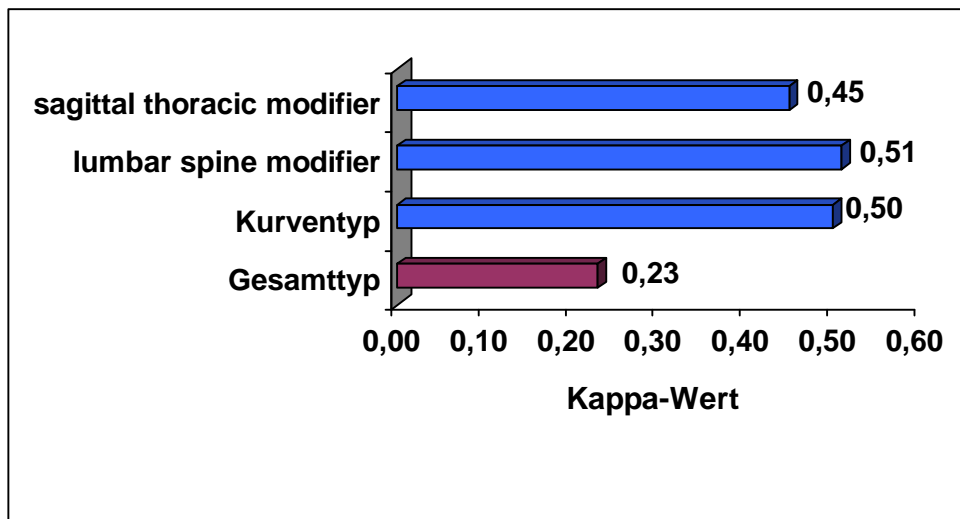


Abbildung 20: Vergleich Lenke-Einzelschritte und Gesamtklassifikation

### 4.3.3 Lernkurven

Die unterschiedliche Entwicklung der Interobserver-Zuverlässigkeit beider Skolioseklassifikationen im Untersuchungsverlauf wird in Abbildung 21 dargestellt. Zu Beginn der Untersuchung lagen die Kappa-Koeffizienten für beide Klassifikationen auf niedrigem Niveau nahe beieinander. Im weiteren Verlauf entfernten sich die Werte zunehmend voneinander. Für die King-Klassifikation war eine stetige und ausgeprägte Steigung übereinstimmender Ergebnisse bei allen drei Untersuchern zu beobachten. Bei der Lenke-Klassifikation verlief die Kurve flacher und ondulierend. Dies galt ebenso für die einzelnen Teilschritte der Lenke-Klassifikation (siehe auch Abbildung 19, Seite 30).

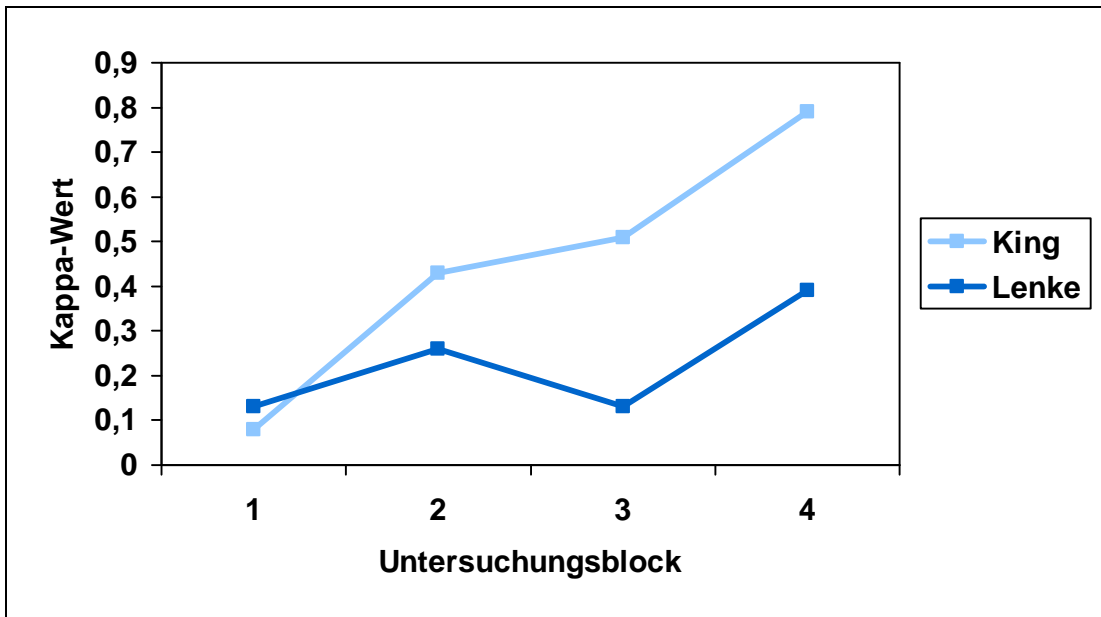


Abbildung 21:  $\kappa$ -Statistik beider Klassifikationen im Untersuchungsverlauf

Die folgenden beiden Abbildungen geben für beide Klassifikationen die Entwicklung aufgeschlüsselt nach der Anzahl übereinstimmender Untersucher wieder (Abbildung 22 und 23).

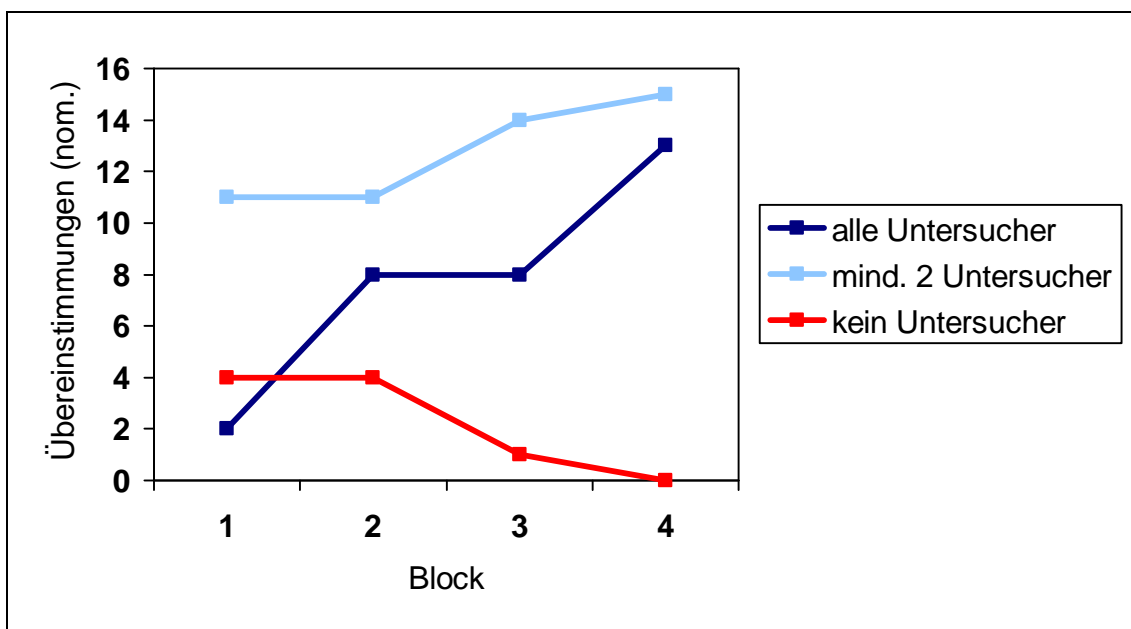


Abbildung 22: Lernkurven King-Klassifikation



Werden nur Übereinstimmungen aller Untersucher betrachtet, zeigt sich für die Einteilung nach King eine steile Trendlinie (dunkelblaue Kurve in Abbildung 22) mit Beginn auf niedrigem Niveau (zwei Fälle). Die hellblaue Kurve stellt eine Übereinstimmung von mindestens zwei Untersuchern dar. Im ersten Auswertungsblock wurden so nominal mehr als zwei Drittel der Patienten identisch klassifiziert. Im letzten Untersuchungsblock gab es keinen Patienten, den alle Untersucher unterschiedlich beurteilten (rote Kurve).

Die fehlenden Übereinstimmungen reduzierten sich auch für die Lenke-Klassifikation (Abbildung 23) im Untersuchungsverlauf von sechs auf zwei (rote Kurve). Auch im letzten Auswertungsblock konnte keine hundertprozentige Übereinstimmung von zwei oder drei Untersuchern erreicht werden (hellblaue Kurvenlinie). Die Anzahl der Übereinstimmungen war insgesamt geringer als bei der Bewertung nach King. Die entsprechende Lernkurve verläuft flacher als die der King-Klassifikation. Auf niedrigem Niveau stieg die Rate an Übereinstimmungen aller Untersucher von Block 1 bis Block 4 unstetig an (dunkelblaue Kurve 23).

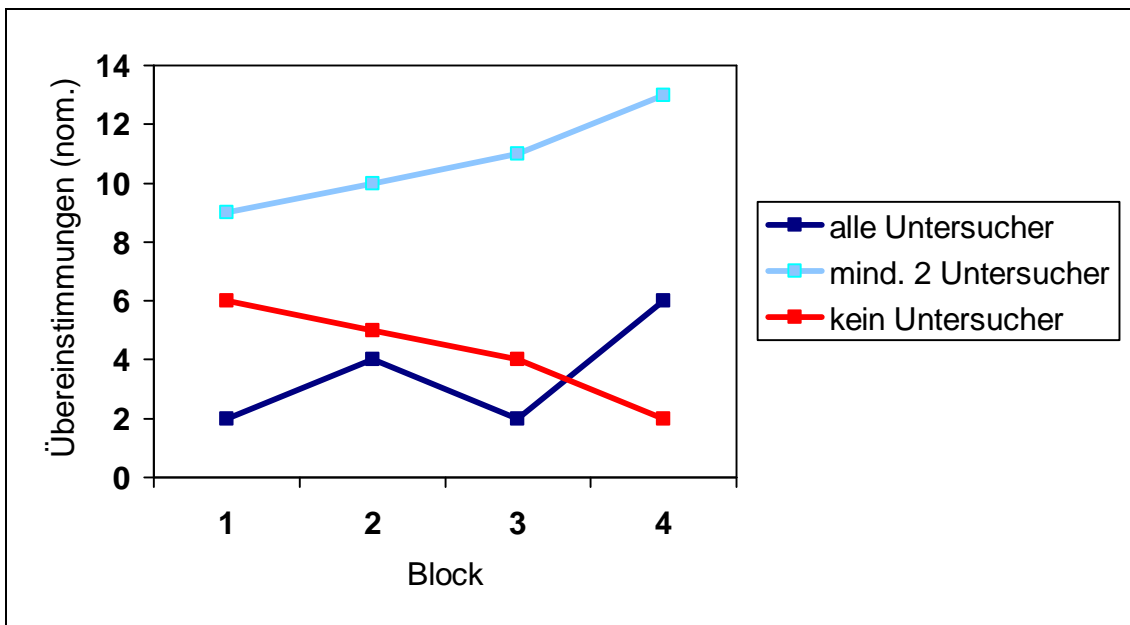


Abbildung 23: Lernkurven Lenke-Klassifikation

#### 4.3.4 Interobserver-Reliabilität verschiedener Untersucherpaare

Der Interobserver-Vergleich wurde auch separat für die Übereinstimmungen verschiedener Untersucherpaare durchgeführt (Abbildung 24). Damit sollte die Abhängigkeit der Reliabilität des Klassifikationsergebnisses vom Ausbildungsstand der Untersucher beurteilt werden.

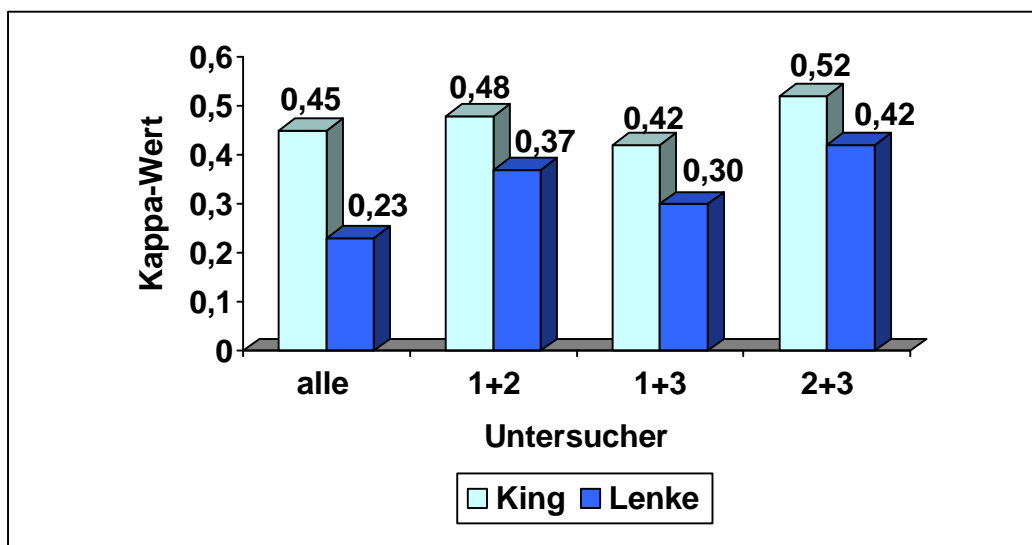


Abbildung 24: Kappa-Statistik verschiedener Untersucherpaare

Die Kappa-Statistik der King-Klassifikation zeigte für die Untersucherpaarungen ähnliche Werte wie bei gemeinsamer Betrachtung aller Untersucher. Die höchste Reliabilität wurde zwischen Untersucher 2 und Untersucher 3 erzielt ( $\kappa=0,52$ ).

Insgesamt niedriger lagen die entsprechenden Werte bei der Lenke-Klassifikation. Die Zuverlässigkeit der Untersucherpaarungen lag über derjenigen für alle Untersucher. Mit einem Kappa-Koeffizienten von 0,42 (+0,19 im Vergleich zu allen drei Untersuchern) erreichten auch in dieser Klassifikation Untersucher 2 und Untersucher 3 die meisten Übereinstimmungen.

Für die komplexere Beurteilung nach Lenke werden im Folgenden diese Ergebnisse nochmals aufgeschlüsselt (Abbildung 25).

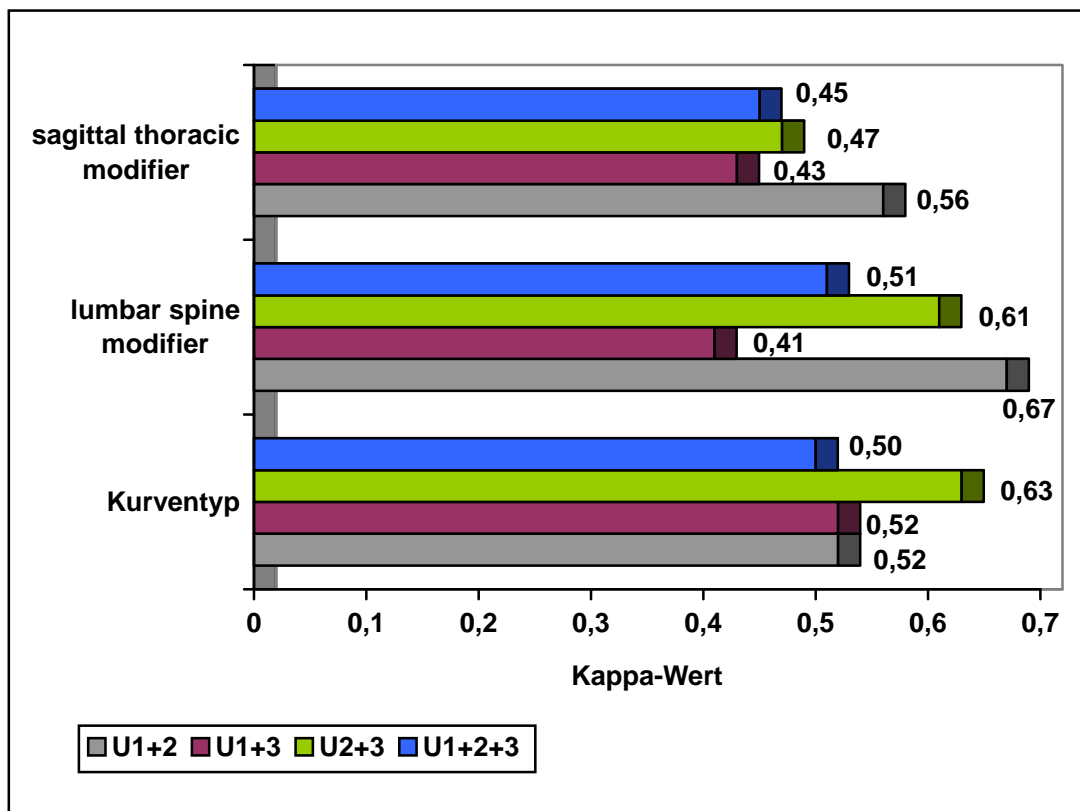


Abbildung 25: Interobserver-Vergleich Lenke-Teilschritte für verschiedene Untersucherpaare

In den einzelnen Teilschritten erreichten die verschiedenen Untersucherpaare zumeist höhere Kappa-Koeffizienten als alle Untersucher gemeinsam. Lediglich Untersucherpaar 1+3 zeigte in zwei Teilschritten niedrigere Werte. Im Vergleich zum Gesamttyp der Lenke-Klassifikation waren die Kappa-Koeffizienten ebenfalls höher (siehe auch Diskussion Seite 45).

Sehr unterschiedlich stellten sich die Lernkurven für die Untersucherpaare im Untersuchungsverlauf dar. Die steilste Entwicklung war für die Übereinstimmungen zwischen Untersucher 2 und Untersucher 3 festzustellen (Abbildung 26, grüne Kurve). Flacher verläuft die Lernkurve für Untersucher 1 und Untersucher 3 (lila Kurve). Bei der Paarung Untersucher 1 und Untersucher 2 ist keine Steigung der Lernkurve zu objektivieren (graue Kurve).

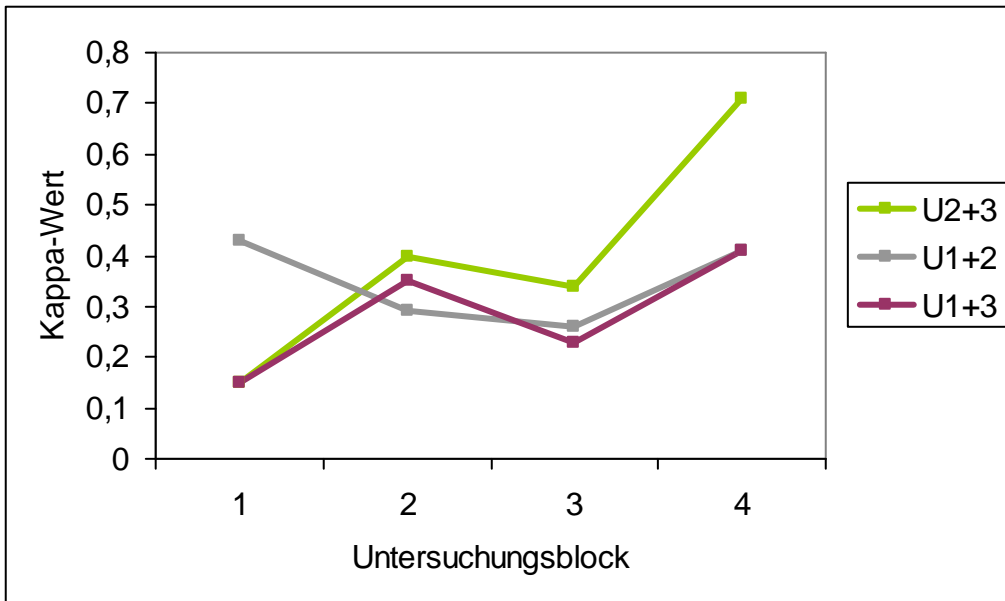


Abbildung 26: Lernkurven Untersucherpaare für die Lenke-Klassifikation

## **5. Diskussion**

### **5.1 Allgemeines**

Der Stellenwert guter klinischer Klassifikationen ist unumstritten. Sie sind ein wichtiges Instrument zur Standardisierung medizinischer Sachverhalte. Sie ermöglichen bei entsprechender Akzeptanz und Verbreitung im klinischen Alltag sowohl die Erarbeitung diagnostischer und therapeutischer Richtlinien als auch die Schaffung einer einheitlichen Basis für wissenschaftliche Untersuchungen.

Winter weist in seinen Ausführungen auf die Schwierigkeiten bei der Entwicklung solcher Klassifikationen für die Orthopädie hin. Die Vielfalt und Komplexität orthopädischer Sachverhalte erschwert die sinnvolle Sortierung. Der medizinische Kenntnisstand und damit die Aktualität entsprechender Klassifikationen unterliegen permanenten Veränderungen durch wissenschaftliche Forschung. Zwangsläufig stößt deshalb die Entwicklung möglichst vollständiger, zeitgerechter Klassifikationen an ihre Grenzen (Winter 1999).

Angesichts dieser Tatsachen erscheint es verständlich, dass die vorliegende Studie bei der kritischen Prüfung zweier Skolioseklassifikationen Defizite aufzeigte. Die reine Betrachtung der Statistik für die King-Klassifikation und die von Lenke vorgeschlagene Einteilung mag zunächst ernüchternd wirken und enttäuschen. Bei der Suche nach Gründen für diese Ergebnisse sollen im Folgenden deshalb, im Bemühen um eine möglichst sinnvolle und realistische Bewertung, verschiedene weitere Aspekte und Sachverhalte betrachtet werden. Neben Reproduzierbarkeit und Zuverlässigkeit der Klassifikationen spielen weitere Kriterien eine Rolle bei der Beurteilung und Prüfung der beiden Klassifikationen. Die vollständige Abbildung des inhaltlichen Zusammenhangs, die Möglichkeit der Entwicklung therapeutischer Empfehlungen, einfache und leicht erlernbare Anwendung sowie klinische Praktikabilität sind unter anderem

von Lenke formulierte Anforderungen an eine Klassifikation (Lenke et al. 2001). Lonstein stellte in seinen Ausführungen zu einer Publikation von Richards et al. fest, dass die Akzeptanz jeder Klassifikation auch von diesen Voraussetzungen abhängig ist (Lonstein 2003, Richards et al. 2003).

Aufgrund dessen wurde der Schaffung möglichst realitätsnaher, dem klinischen Routineablauf entsprechender Untersuchungsbedingungen in dieser Arbeit besondere Beachtung geschenkt. Ganz bewusst wurden Untersucher unterschiedlichen Ausbildungsstandes gewählt. Der Einfluss dieser Bedingungen auf das Untersuchungsergebnis wird unter anderem ausführlicher in der Beurteilung der Lenke-Klassifikation diskutiert (siehe Seite 43 und 44). Die Beurteilung von Lerneffekten im Untersuchungsverlauf und die Prüfung des Einflusses klinischer Erfahrung auf das Klassifikationsergebnis stellten einen Schwerpunkt dieser Studie dar.

## **5.2 King-Klassifikation**

Die 1983 von King et al. vorgestellte Klassifikation idiopathischer Skoliosen wurde als Entscheidungshilfe für die Planung der operativen Therapie entwickelt und hat weltweit eine große Verbreitung und Akzeptanz gefunden. Cummings et al. prüften 1998 die Reproduzierbarkeit und Reliabilität dieser Einteilung. Mit einer medianen Intraobserver-Zuverlässigkeit von  $\kappa=0,64$  (Streuung 0,44-0,72) lagen die Werte dieser Studie in ähnlichen Größenordnungen wie die der vorliegenden Untersuchung ( $\kappa=0,53$ -0,74). Auch für den Interobserver-Vergleich waren die Ergebnisse fast identisch (Cummings et al.:  $\kappa=0,44$ ; diese Studie:  $\kappa=0,45$ ). Ebenfalls 1998 publizierten Lenke et al. eine Untersuchung zur Zuverlässigkeit der King-Klassifikation mit gering niedrigeren Werten (Intraobserver-Kappa=0,62, Interobserver-Kappa=0,40). Somit bestätigen sich für die King-Klassifikation entsprechend der Kriterien von Svanholm et al. von 1989 eine lediglich mäßige Intraobserver- und eine schlechte Interobserver-Zuverlässigkeit.

Das Studiendesign von Cummings et al. war denen der vorliegenden Untersuchung vergleichbar. Die beteiligten Untersucher hatten unterschiedliche klinische Erfahrungen zum Teil ohne Bezug zur Skoliose-therapie. Die verwendeten Röntgenbilder enthielten keinerlei Winkelangaben oder andere Markierungen. In der Studie von Lenke et al. dagegen waren die notwendigen Winkel im Vorfeld einheitlich vermessen worden und auf den Röntgenbildern eingetragen. Die Untersuchergruppe bestand nur aus aktiven Skoliose-Chirurgen, wenn auch mit unterschiedlicher operativer Erfahrung.

Trotz dieser erheblichen Unterschiede kamen alle drei Studien zu ähnlichen Ergebnissen. Der Einfluss der klinischen bzw. operativen Erfahrung scheint für die King-Klassifikation demnach nicht von entscheidender Bedeutung zu sein. Diese Schlussfolgerung wird durch die hier dargelegten Ergebnisse der separaten Betrachtung zur Zuverlässigkeit verschiedener Untersucherpaare gestützt. Entgegen der nahe liegenden Annahme wurde die geringste Fehleranfälligkeit nicht zwischen den erfahreneren Untersuchern sondern zwischen Untersucher 2 und Untersucher 3 erreicht (siehe auch Seite 34). Die Kappa-Werte differierten maximal um 0,07 im Vergleich zur Interobserver-Reliabilität aller Untersucher.

Neben der geringen Zuverlässigkeit wurden auch die inhaltlichen Defizite der King-Klassifikation deutlich. Vier nicht klassifizierbare Skliosetypen entsprachen 6,7% des gesamten Patientengutes. Damit wird dieser, in früheren Publikationen mehrfach kritisierte Aspekt der Klassifikation, erneut belegt (Lenke et al 1998 und 2001, Richards et al. 2003). Mit der Intention die genannten Nachteile zu reduzieren und angesichts deutlich vielfältigerer operativer Instrumentationsmöglichkeiten stellten Lenke et al. 2001 eine neue Klassifikation vor.

### 5.3 Lenke-Klassifikation

Bisher existieren nur wenige Arbeiten zur dieser neueren Skolioseklassifikation. Größtenteils stammen die Untersuchungen zur Zuverlässigkeit der Klassifikation aus dem amerikanischen Raum (Lenke et al. 2001, Richards et al. 2003, Puno et al. 2003) Aus Europa ist eine Arbeit von Ogon et al. 2002 veröffentlicht. Andere Veröffentlichungen befassen sich mit dem Zusammenhang von Lenke-Typ und operativer Therapie (Lenke et al. 2001, Lenke et al. 2002, Puno et al. 2003). In Deutschland ist die Klassifikation noch nicht sehr verbreitet. Bis zum Beginn dieser Studie war keine Arbeit zur Anwendung und Prüfung der Lenke-Klassifikation unter hiesigen klinischen Bedingungen veröffentlicht.

Die Ergebnisse der vorliegenden Untersuchung zeigen große Unterschiede zu den meisten der bisher veröffentlichten Arbeiten zur Reliabilität. Die Kappa-Werte für den Intra- und auch den Interobserver-Vergleich der Lenke-Klassifikation liegen unter den Werten anderer Untersuchungen.

Am deutlichsten zeigt sich das im Vergleich zu den Ergebnissen der Publikation von Lenke (Lenke et al. 2001). Die dort ermittelte Fehleranfälligkeit sowohl untersucherspezifisch als auch zwischen den Untersuchern ist so gering, dass selbst unter Anwendung der strengeren Svanholm-Kriterien (Svanholm et al. 1989) eine gute bis exzellente Zuverlässigkeit für die einzelnen Teilschritte erreicht wurde. ( $\kappa > 0,75$ ). Die entsprechenden Werte der hier vorliegenden Untersuchung liegen dagegen für den Interobserver-Vergleich weitaus niedriger ( $\kappa = 0,45-0,51$ ). Für den sagittal thoracic modifier beträgt diese Differenz sogar mehr als 100 %. Die Zuverlässigkeit des Lenke-Gesamttypes kann nicht verglichen werden, da in Lenke's Publikation kein Kappa-Wert für die Gesamtklassifikation angegeben wurde.

Nicht ganz so ausgeprägt aber dennoch deutlich sind auch die Unterschiede der Ergebnisse verglichen zur Publikation von Ogon et al. 2002. Der



Interobserver-Kappa wurde dort mit  $\kappa=0,69-0,91$  in den Teilschritten angegeben. Der angegebene Kappa-Wert von 0,62 für die Gesamtklassifikation ist im Vergleich zur vorliegenden Arbeit sogar fast dreifach so hoch.

#### **5.4. Einfluss der Cobb-Winkelmessung**

Einen entscheidenden Grund für die hier dargestellten schlechteren Ergebnisse der Lenke-Klassifikation im Vergleich zu anderen Untersuchungen stellt der Studienaufbau dar.

Eine Erklärung für die Diskrepanz der Resultate ist dabei in der Auswertungsgrundlage zu finden. Diese unterscheidet sich in den oben erwähnten früheren Untersuchungen zum Teil deutlich von dieser Arbeit. In den Untersuchungen von Lenke et al. (2001) und Ogon et al. (2002) wurden den Untersuchern markierte Röntgenbilder präsentiert. Es waren sowohl wichtige Bezugspunkte bzw. -linien eingezeichnet als auch die Cobb-Winkel bereits vermessen und angegeben. Lediglich Richards et al. (2003) untersuchten nicht vormarkierte Röntgenbilder und erreichten erwartungsgemäß niedrigere Kappa-Werte in der Gesamtklassifikation.

Die Wahl der Bezugspunkte, das Einzeichnen der Hilfslinien und die Winkelmessung sind sehr individuell geprägte Vorgänge. Trotz exakter Regeln zur Festlegung und Durchführung der Markierungen und Messung sind Abweichungen unvermeidlich. Das trifft sowohl für Wiederholungsmessungen eines Einzelnen als auch für den Vergleich mehrerer Untersucher zu.

Morrissy et al. beschäftigten sich in ihrer Arbeit ausführlich mit der intrinsischen Fehleranfälligkeit der Cobb-Winkelmessung bei Patienten mit Skoliosen (Morrissy et al. 1990). Dabei wurde bei individueller Wahl der Bezugspunkte eine Intraobserver-Variabilität von  $4,9^\circ$  ermittelt. Für den Interobserver-Vergleich erhöhte sich diese Zahl bis auf  $7,2^\circ$ . In 4,2 % aller Cobb-Messungen

wichen die Untersucher mehr als 5° in den Ergebnissen voneinander ab. Waren die Endwirbelkörper dagegen im Vorfeld bereits festgelegt, reduzierten sich Abweichungen dieser Größenordnung auf unter 1%. Die Autoren stellten außerdem einen Einfluss des verwendeten Winkelmessers auf die Höhe des Messfehlers fest.

Diese Resultate illustrieren eindrucksvoll, wie entscheidend die korrekte Festlegung der Bezugspunkte und die Winkelmessung für die Klassifikation sind. Insbesondere Skoliosen, die in ihrer Ausprägung nahe der definierten Grenzen von Klassifikationstypen liegen, sind für diesen Fehler anfällig. Messunterschiede in oben genannter Höhe können bereits zu verschiedenen Typzuordnungen führen. Verschärft wird das Problem durch die unterschiedliche Qualität der Röntgenbilder. Dieses Kriterium stellt eine weitere Fehlerquelle dar. Insbesondere die seitlichen Wirbelsäulenaufnahmen erschweren durch Überlagerungen die Festlegung benötigter Bezugspunkte und damit die korrekte Vermessung. Belegt wird diese Aussage in der vorliegenden Studie durch die niedrigeren Kappa-Werten für den sagittal thoracic modifier. Dieser Teilschritt der Lenke-Klassifikation war der unzuverlässigste von allen. Bezüglich der Bedeutung einer individuellen Vermessung auf das Klassifikationsergebnis bestätigen sich damit die von Richards et al. 2003 formulierten Resultate und Schlussfolgerungen.

Eine Elimination dieser Fehlerquellen durch Winkelangaben auf den Röntgenbildern (z.B. bei Lenke et al. 2001 und Ogon et al. 2002) ist für die Prüfung der klinischen Anwendbarkeit kritisch zu betrachten. In der klinischen Routine stehen dem Untersucher im Normalfall keine markierten Röntgenunterlagen zur Verfügung. Zudem ist, wie im Falle der Lenke-Klassifikation, bei einer Vielzahl zu messender Winkel von einer Kumulation des Fehlers auszugehen. Der Wegfall dieses zentralen Bestandteils erlaubt deshalb zwar eine gute Prüfung des Verständnisses für die Klassifikation, nicht jedoch der Zuverlässigkeit im klinischen Routinebetrieb.

## 5.5 Einfluss der Erfahrung unterschiedlicher Untersucher

Ein weiterer Aspekt ist die Auswahl der Untersucher. In den früheren Studien waren entweder klinisch und/oder mit der Lenke-Klassifikation erfahrene Untersucher beteiligt. (Lenke et al. 2001, Ogon et al. 2002, Richards et al. 2003). Bei Lenke et al. wurden die Werte einer Gruppe von Initiatoren der neuen Klassifikation mit denen einer Gruppe randomisiert ausgewählter Skoliose-Chirurgen verglichen. In den Studien von Ogon et al. (2002) und Richards et al. (2003) waren die Untersucher klinisch unterschiedlich erfahren, hatten aber bereits die neue Lenke-Klassifikation regelmäßig angewandt. Diese Untersuchungsbedingungen sind jedoch nicht für den Alltag orthopädischer Kliniken in Deutschland repräsentativ. Es wurde deshalb versucht, für die vorliegende Studie realitätsnähere Bedingungen zu schaffen. Die beteiligten Untersucher waren alle unerfahren in der Anwendung der Lenke-Klassifikation und hatten zusätzlich einen unterschiedlichen Ausbildungsstand in Bezug auf die Wirbelsäulenchirurgie.

Die Auswirkungen dieses Kriteriums auf das Klassifikationsergebnis sind nicht mit denen bereits eingezeichneter Bezugspunkte und Winkel zu vergleichen. Dennoch ist im Gegensatz zur King-Klassifikation für die Zuverlässigkeit der Lenke-Klassifikation ein Einfluss durch die Erfahrung der Untersucher anzunehmen. Der Interobserver-Gesamtkappa-Wert ist trotz vergleichbaren Untersuchungsmaterials (unmarkierte Röntgenbilder) in der Arbeit von Richards et al. von 2003 mit 0,50 noch mehr als doppelt so hoch als in dieser Studie.

Erwartungsgemäß betrifft dies insbesondere die Kombination von wenig klinischer Erfahrung mit Unerfahrenheit in der Anwendung der Klassifikation. Der Intraobserver-Kappa für den Untersucher 3 ( $\kappa=0,30$ ) lag weit unter dem der beiden anderen Untersucher ( $\kappa=0,51$  bzw.  $0,57$ ). Er ist er nur halb so hoch wie der bei Richards et al. (2003) angegebene Durchschnittswert von 0,60 für die Gesamtklassifikation.

Interessanterweise korreliert jedoch das Ausmaß klinisch-operativer Erfahrung nicht zwangsläufig mit der Güte des Klassifikationsergebnisses. In dieser Untersuchung erreichte der Untersucher 2 und nicht der klinisch erfahrenere Untersucher 1 bei beiden Klassifikationen die höchsten Intraobserver-Werte und damit die beste Reproduzierbarkeit. Diese Aussage wird auch durch die Ergebnisse des Interobserver-Vergleiches belegt. Die Untersucherpaarung Untersucher 2 und Untersucher 3 erreichte den höchsten Gesamtkappa-Wert. Untersucher 1 und Untersucher 2 dagegen zeigten im entsprechenden Vergleich die geringste Zuverlässigkeit. Auch die Lernkurven der Untersucherpaare spiegeln diese Tatsache eindrücklich wieder (siehe Ergebnisse Seite 37). Eine Ausnahme stellt lediglich die Bewertung des sagittalen Profils dar. In diesem Lenke-Teilschritt war die Übereinstimmung zwischen Untersucher 2 und Untersucher 3 geringer als die der anderen Untersucherpaare (siehe Ergebnisse Seite 35).

Es sind verschiedene Erklärungen für dieses Resultat möglich. Einerseits ist das Niveau, auf dem der klinisch Unerfahrene beginnt deutlich niedriger als das des Erfahrenen. Die Lernkurve von Untersucher 1 und Untersucher 2 (Abbildung 26, Seite 36) beginnt im ersten Untersuchungsblock bei einem knapp dreifach höheren Wert verglichen mit den beiden anderen Untersucherpaaren (U2/U3 und U1/U3). Diese wiederum haben exakt denselben Ausgangswert. Daraus lässt sich vermuten, dass unabhängig vom zweiten Untersucher zumindest zu Beginn die Unerfahrenheit des Untersuchers 3 entscheidend für die Zuverlässigkeit des Ergebnisses ist. Eine zweite mögliche Erklärung wurde von Lenke 2002 in seinen Ausführungen zur Publikation von Ogon et al. (2002) beschrieben. Für zuverlässige Resultate in der Lenke-Typisierung ist ein striktes Festhalten an den Klassifikationsrichtlinien nötig. Erfahrene Kliniker der Skoliotherapie neigen eher dazu, klinisch wichtige röntgenologische Zusatzinformationen unbewusst in die Bewertung einfließen zu lassen. Diese Informationen (z.B. T1-Verkipfung oder Schultergeradstand) sind jedoch nicht immer auch relevant für die Festlegung des Lenke-Types. Dadurch kann es zu einer ungewollten

Fehlklassifikation kommen. Diese Fehlerquelle gibt es für den klinisch wenig Erfahrenen kaum (Lenke 2002).

## 5.6 Einfluss Kappa-Statistik

Unabhängig vom Studiendesign wird in allen Arbeiten zur Reliabilität der King- und Lenke-Klassifikation die Kappa-Statistik angewandt.

In den meisten klinisch angewandten Klassifikationen ist der Untersucherfehler von Bedeutung. Zufällige Übereinstimmungen als eine Fehlerkomponente können in Reliabilitäts-Studien mit der Kappa-Statistik berücksichtigt werden. Das ist unbestritten ein großer Vorteil dieser statistischen Methode. Je höher die Wahrscheinlichkeit einer oder mehrerer Untersucher ist, zufallsbedingt zum selben Ergebnis zu kommen, umso mehr wird das Resultat beobachteter Übereinstimmungen korrigiert.

Ein Faktor, der diese Wahrscheinlichkeit beeinflusst, ist u.a. die Komplexität der geprüften Klassifikation. Bei 42 möglichen Lenke-Typen ist ein zufällig übereinstimmendes Klassifikationsergebnis zwischen allen Untersuchern so gut wie nicht zu erwarten ( $1/42 \cdot 1/42 \cdot 1/42 = 3,21E-07$ ). Bei lediglich 5 verschiedenen King-Typen dagegen ist diese Wahrscheinlichkeit größer ( $1/5 \cdot 1/5 \cdot 1/5 = 0,008$ ). Deshalb führte die Anwendung der Kappa-Statistik für die King-Klassifikation in dieser Untersuchung zu stärkeren Korrekturen. Ein weiteres Beispiel für den Einfluss der Kappa-Statistik auf die Wertung von Ergebnissen zeigt sich in der Lenke-Klassifikation. Die Ermittlung des Gesamttypes ist komplexer als die der Teilschritte. Trotzdem lag der Kappa-Wert der Gesamtklassifikation unter denen der Teilschritte. Dieser Unterschied ist also nicht zufällig bedingt. Die Zuverlässigkeit der gesamten Lenke-Klassifikation ist demnach erwartungsgemäß tatsächlich geringer als die der Teilschritte (siehe auch Ergebnisse, Seite 31). Auch die Anzahl der beteiligten Untersucher ist bei Ermittlung des Kappa-Wertes von Bedeutung. Es ist wahrscheinlicher, dass zwei Untersucher sowohl real als auch zufällig übereinstimmen als drei. Die

Kappa-Koeffizienten für die Untersucherpaare sind dementsprechend zumeist höher als für alle Untersucher (siehe auch Ergebnisteil Seite 35).

Die Kappa-Statistik beinhaltet auch einige methodische Fehlerquellen. Richards et al. wies in diesem Zusammenhang auf die Bedeutung einer unausgeglichene Verteilung vorkommender Kategorien hin. Ist eine Kategorie übermäßig häufig in den Ergebnissen vertreten, ist die Wahrscheinlichkeit zufälliger Übereinstimmungen in dieser Kategorie hoch. Diese hohe Zufallswahrscheinlichkeit wird aber nicht durch die Klassifikation als solche bedingt, sondern durch die ungleiche Verteilung. Der resultierende kleine Kappa-Koeffizient suggeriert dann eine falsch niedrige Zuverlässigkeit des Verfahrens (Richards et al. 2003).

Thomson et al. untersuchten die Signifikanz der Anzahl beteiligter Untersucher (Thomson et al. 2002). Es wurden für die Klassifikationen von Weber und Lauge-Hansen (Weber 1972, Hansen 1942) Kappa-Statistiken mit einer steigenden Anzahl von Untersuchern erhoben. Bei einer geringen Anzahl streuten die Kappa-Koeffizienten erheblich (0,27-0,90 in der Weber-Klassifikation bei 3 Untersuchern). Mit steigender Anzahl an Untersuchern stabilisierte sich der Kappa-Wert. Bei kleinen Untersuchergruppen stellten Thomson et al. deshalb die Aussagekraft der Kappa-Statistik zur alleinigen Beurteilung der Zuverlässigkeit in Frage. Eine Überprüfung der Skolioseklassifikationen im Rahmen einer weiteren Untersuchung mit weitaus mehr Untersuchern wäre deshalb wünschenswert.

Angesichts dieser möglichen methodischen Fehler müssen alle Studien, die die Kappa-Statistik verwenden, entsprechend vorsichtig gewertet werden. Das gilt auch für die hier dargestellten Ergebnisse. Hinweise zur Problematik der verwendeten Klassifikationen oder Tendenzen eines Lerneffektes lassen sich jedoch trotz möglicher statistischer Defizite aus den Resultaten dieser Studie ableiten.

## **5.7 Zuverlässigkeit anderer orthopädischer Klassifikationen**

In der Orthopädie haben viele Klassifikationen weite Verbreitung gefunden und sind Basis wissenschaftlicher Forschung. Die Reliabilität dieser Klassifikationen ist aber trotz hoher klinischer Akzeptanz nicht immer gut. Die Weber-Klassifikation erreichte in der Untersuchung von Thomsen et al. (1991) beispielsweise nur eine mäßige Interobserver-Zuverlässigkeit ( $\kappa=0,58$ ) entsprechend der Svanholm-Kriterien (Svanholm et al. 1989). Die Neer-Klassifikation für proximale Humerusfrakturen sowie das AO-Klassifikationssystem für distale Radiusfrakturen schnitten in entsprechenden Studien noch schlechter ab (Sidor et al. 1993, Illarramendi et al. 1998). Sie zeigten ebenso wie die Magerl-Klassifikation für Wirbelfrakturen eine schlechte Zuverlässigkeit (Oner et al. 2002).

Die Zuverlässigkeit der hier getesteten Skolioseklassifikationen differiert demnach nicht wesentlich verglichen mit diesen etablierten Klassifikationen. Gleichzeitig verdeutlicht diese Tatsache die Diskrepanz zwischen Anforderung an klinische Klassifikationen und Umsetzung. Prinzipiell sollte nur eine zuverlässige Klassifikation auch klinisch angewendet werden. Das entspricht, wie die obigen Beispiele belegen, nicht immer der Realität.

## **5.8 Schlussfolgerungen und Ausblick**

Die neuere Lenke-Klassifikation ist der etablierten Einteilung von King in der inhaltlichen Erfassung der verschiedenen Skolioseformen deutlich überlegen. Insbesondere im Hinblick auf die in letzten Jahren stetig verbesserten operativen Optionen ist dies bedeutend (Lenke et al. 2001, Lenke et al. 2002, Puno et al. 2003). In diesem Zusammenhang kann die King-Klassifikation als nicht mehr zeitgerecht betrachtet werden. Sie sollte deshalb die ältere King-Klassifikation im klinischen Alltag und insbesondere bei wissenschaftlichen Arbeiten ablösen.

Beide Klassifikationen vernachlässigen die ätiologisch und klinisch relevante Rotation der Wirbelkörper. Für die Beurteilung dieser Eigenschaft der Skoliose sollten deshalb ein zusätzliches Kriterium eingearbeitet werden oder eine Kombination mit Klassifikationen erfolgen, die diese Komponente berücksichtigen (z.B. Nash and Moe 1969).

Die Lenke-Klassifikation ist deutlich komplexer als die King-Klassifikation. Es bedarf einer längeren, intensiven und regelmäßigen Anwendung zum Erlernen der Klassifikation. Zudem beeinflusst der Ausbildungsstand der Untersucher das Klassifikationsergebnis und die Lernkurve. Die Erfahrung der Untersucher spielt bei der Einteilung nach King dagegen kaum eine Rolle. Durch eine Vielzahl festzulegender Bezugspunkte und zu messender Winkel ist die Lenke-Klassifikation deutlich fehleranfälliger als die ältere Klassifikation. Unter realitätsnahen Bedingungen erreicht sie eine geringere Zuverlässigkeit als die Einteilung von King und anderer gängiger orthopädischer Klassifikationen (siehe auch Seite 47).

Trotz vorgenannter Schwierigkeiten erscheint eine Integration der Lenke-Klassifikation in den klinischen Routineablauf sinnvoll. Jedoch sind zum jetzigen Zeitpunkt die Bedingungen dafür schwierig. Sind in der Zukunft etablierte Verfahren verfügbar, die eine voll- oder teilweise automatisierte Vermessung erlauben, können wichtige Fehlerquellen eliminiert werden. Ein solcher Algorithmus wurde zum Beispiel von Stokes und Aronsson 2002 vorgestellt. Vielleicht sind auch durch neue digitale Röntgentechnik mit entsprechender Vermessungssoftware eine bessere Bildqualität und eine Objektivierung der Messung möglich. Dies sollte in weiterführenden Untersuchungen geklärt werden.

Liegen entsprechende technische Voraussetzungen vor, ist es sinnvoll, die Lenke-Klassifikation im klinischen Alltag einzusetzen. Die regelmäßige Anwendung führt zu routiniertem, sicherem Umgang mit der Klassifikation. Insbesondere sollte bereits zu Beginn der klinischen Ausbildung mit der



konsequenter Heranführung an die Klassifikation begonnen werden (siehe auch Diskussion Seite 44).

Mit diesen Maßnahmen kann die Praktikabilität der komplexen Lenke-Klassifikation deutlich verbessert werden. Dadurch erhöhen sich sowohl die Zuverlässigkeit als auch die Chancen auf eine weit verbreitete klinische Anwendung und Akzeptanz. Unter diesen Voraussetzungen ist die Klassifikation von Lenke unter den Gesichtspunkten der modernen Skoliosebehandlung geeigneter als die ältere Einteilung von King.

## 6. Zusammenfassung

Für effektives Arbeiten, zur klinischen Entscheidungsfindung und für wissenschaftliche Forschung haben medizinische Klassifikationen eine große Bedeutung. Ziel dieser Arbeit war es deshalb, die neuere Lenke-Klassifikation für idiopathische Skoliosen mit der älteren etablierten Einteilung von King hinsichtlich Zuverlässigkeit und Praktikabilität unter realitätsnahen Bedingungen zu prüfen.

Drei Untersucher, alle unerfahren mit der Lenke-Klassifikation und mit unterschiedlicher klinischer Erfahrung in der Skoliotherapie, klassifizierten unmarkierte Röntgenbilder von 60 Patienten mit idiopathischer Skoliose. Die Blindstudie wurde in 4 zeitlich getrennten Untersuchungsblöcken durchgeführt. Die Auswertung erfolgte mit der Multi-Rater-Kappa-Statistik und dem Student-t-Test für unverbundene Stichproben. Beide Klassifikationen zeigten eine geringe bis mäßige Intraobserver- und eine geringe Interobserver-Reliabilität (Interobserver- $\kappa$ : Lenke=0,23, King=0,45). Die Winkelmessung und die Erfahrung der Untersucher haben insbesondere bei der Lenke-Einteilung einen Einfluss auf die Zuverlässigkeit. Die Lernkurve verläuft für diese Klassifikation flacher als bei der Einteilung von King. Die Lernkurve von Untersuchern geringerer Erfahrung steigt steiler an und die Reliabilität ist höher.

Mit der Lenke-Klassifikation können idiopathische Skoliosen umfassender und detaillierter beschrieben werden als mit der älteren King-Klassifikation. Damit wird diese Klassifikation den neueren Therapieoptionen besser gerecht. Trotz geringerer Zuverlässigkeit und Praktikabilität sollte sie deshalb wissenschaftlich und klinisch eingesetzt werden. Unter Reduzierung wichtiger Fehlerquellen mittels teil- oder vollautomatisierten Vermessung und regelmäßiger Anwendung ist eine Steigerung der Zuverlässigkeit zu erwarten und damit ein Einsatz unter klinischen Routinebedingungen denkbar.

## 7. Literaturverzeichnis

Cobb J.R. (1948)

Outline for the study of scoliosis 261-275.

in: Instructional Course Lectures. The American Academy of Orthopaedic Surgeons. Vol. 5

Ann Arbor, J.W. Edwards

Cummings R.J., Loveless E.A., Campbell J., Samelson S., Mazur J.M. (1998)

Interobserver reliability and intraobserver reproducibility of the system of King et al. for the classification of adolescent idiopathic scoliosis

J Bone Joint Surg 80-A: 1107-11

Hansen N.L. (1942)

Ankelbrud:genetisk diagnose og reposition.

Dissertation

Copenhagen, Munksgaard

Illarramendi A., González A., Segal E., De Carli P., Maignon G., Galucci G.

(1989)

Evaluation of simplified Frykman and AO classification of fractures of the distal radius – an assessment of interobserver and intraobserver agreement

Int Orthop 22: 111-5

King H.A., Moe J.H., Bradford D.S., Winter R.B. (1983)

The selection of fusion levels in thoracic idiopathic scoliosis

J Bone Joint Surg (Am) 65-A: 1302-13

Krämer K.-L. (1999)

Dokumentation, Voraussetzung für Qualitätsmanagement

Orthopäde 28: 191-200

Krämer K.-L., Maichl F.-P. (1993)

Vorwort, VI-VII

in: Krämer K.-L., Maichl F.-P. unter Mitarbeit von Stock M., Biskup Th.,

Wiedemann J.: Scores, Bewertungsschemata und Klassifikationen in Orthopädie und Traumatologie

Thieme, Stuttgart, New York

Landis J.R., Koch G.G. (1977)

The measurement of observer agreement for categorical data

Biometrics 33: 159-74

Lenke L.G. (2002)

Point of view

Spine 27: 862-3

Lenke L.G., Betz R.R., Bridwell K.H., Clements D.H., Harms J., Lowe T.G., Shufflebarger H.L. (1998)

Intraobserver and interobserver reliability of the classification of thoracic adolescent idiopathic scoliosis  
J Bone Joint Surg 80-A: 1097-106

Lenke L.G., Betz R.R., Harms J., Bridwell K.H., Clements D.H., Lowe T.G., Blanke K. (2001)

Adolescent idiopathic scoliosis: a new classification to determine extent of spinal arthrodesis  
J Bone Joint Surg 83-A : 1169-81

Lenke L.G., Betz R.R., Haheer T.R., Lapp M.A., Merola A.A., Harms J., Shufflebarger H.L. (2001)

Multisurgeon assessment of surgical decision-making in adolescent idiopathic scoliosis – curve classification, operative approach and fusion levels  
Spine 26: 2347-53

Lenke L.G., Betz R.R., Clements D.H., Merola A.A., Haheer T.R., Lowe T.G., Newton P., Bridwell K.H., Blanke K. (2002)

Curve prevalence of a new classification of operative adolescent idiopathic scoliosis – does classification correlate with treatment?  
Spine 27: 604-11

Lonstein J.E. (2003)

Point of view  
Spine 28: 1156-7

Morrissy R.T., Goldsmith G.S., Hall E.C., Kehl Ph.D.D., Cowie G.H. (1990)

Measurement of the Cobb angle on radiographs of patients who have scoliosis  
J Bone Joint Surg 72-A: 320-27

Nash C.L., Moe J.H. (1969)

A study of vertebral rotation  
J Bone Joint Surg 51-A: 223-9

Ogon M., Giesinger K., Behensky H., Wimmer C., Nogler M., Bach C.M., Krismer M. (2002)

Interobserver and intraobserver reliability of Lenke's new scoliosis classification system  
Spine 27: 858-62

Oner F.C., Ramos I.M., Simmermacher R.K., Kingma P.T., Diekerhof C.H., Dhert W.J., Verbout A.J. (2002)

Classification of thoracic and lumbar spine fractures: problems of reproducibility. a study of 53 patients using CT and MRI  
Eur Spine J 11: 235-45

- Puno R.M., An K-C., Puno R.-L., Jacob A., Chung S-S. (2003)  
Treatment recommendations for idiopathic scoliosis – an assessment of the Lenke classification  
Spine 28: 2102-15
- Richards B.S., Sucato D.J., Konigsberg D.E., Ouellet J.A. (2003)  
Comparison of reliability between the Lenke and King classification systems for adolescent idiopathic scoliosis using radiographs that were not premeasured  
Spine 28: 1148-56
- Sidor M.L., Zuckerman J.D., Lyon T., Koval K.B.S., Cuomo F., Schoenberg N. (1993)  
The neer classification system for proximal humeral fractures – an assessment of interobserver reliability and intraobserver reproducibility  
J Bone Joint Surg (Am) 75: 1745-50
- Stokes I.A., Aronsson D.D. (2002)  
Identifying sources of variability in scoliosis classification using a rule-based automated algorithm  
Spine 15: 2801-5
- Svanholm H., Starklint H., Gundersen H.J.G., Fabricius J., Barlebo H., Olsen S. (1989)  
Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic  
APMIS 97: 689-98
- Thomson N.O.B., Overgaard S., Olsen L.H., Hansen H., Nielsen S.T. (1991)  
Observer variation in the radiographic classification of ankle fractures  
J Bone Joint Surg (Br) 73: 676-8
- Weber B.G. (1972)  
Die Verletzungen des oberen Sprunggelenks  
2nd ed. Berne, Verlag Hans Huber
- Winter T. (1999)  
Nomenklatur und Klassifikation in der Orthopädie  
Orthopäde 28: 261-69

## **Danksagung**

Herrn Professor Dr. N. Wülker danke ich herzlich für die Beratung bei der Erstellung dieser Dissertation.

Dr. T. Niemeyer möchte ich für die Vergabe des Themas, seine Betreuung, sein Engagement bei der Durchführung der Untersuchungen sowie seine zahlreichen Anregungen ganz besonders danken.

Mein Dank gilt auch Frau Dr. A. Wolf, die ebenfalls durch ihre Teilnahme und freundliche Unterstützung bei den Untersuchungen diese Dissertation ermöglichte.

Für die Hilfe bei der Recherche und Beschaffung der Patientenunterlagen danke ich den Mitarbeitern des Zentralarchivs der Crona-Kliniken Tübingen.

Bedanken möchte ich mich außerdem herzlich bei Prof. Dr. K. Dietz vom Institut für Medizinische Biometrie für seine freundliche Beratung und Unterstützung bei den statistischen Auswertungen und Berechnungen.

Außerdem möchte ich noch meinen Eltern sowie meinem Mann und Sohn ganz besonders danken, die mit ihrer liebevollen Unterstützung und Geduld die Arbeit an dieser Dissertation ermöglicht haben.

## Lebenslauf

Name, Vorname: Dr. med. dent. KLUBA, Susanne; geb. Schubert  
Geburtsdatum, -ort: 08.05.1973; Halle/ Saale  
Familienstand: verheiratet  
Kinder: Sebastian Kluba, geb. 01. 05. 2002  
Staatsangehörigkeit: deutsch  
Mutter: Dr. Karin Schubert, Zahnärztin  
Vater: Prof. Dr. Dr. Johannes Schubert, Mund-Kiefer-  
Gesichtschirurg

1979 – 1989 Erste Polytechnische Oberschule, Halle-Neustadt  
Mittlere Reife

1989 – 1991 Erweiterte Oberschule, Halle-Neustadt  
Allgemeine Hochschulreife

1991 – 1994 Höhere Berufsfachschule für das Hotel-, Gaststätten-  
und Fremdenverkehrsgewerbe, Saarbrücken;  
staatl. gepr. Assistentin für das Hotel-, Gaststätten-  
und Fremdenverkehrsgewerbe

1994 – 1995 C.C.M., Köln; Projektassistentin

1995 – 1997 Deutsche Lufthansa AG, Stuttgart; Flugbegleiterin

1997 – 2003 Studium der Zahnmedizin an der Eberhard-Karls-  
Universität, Tübingen

seit 1998 Studium der Medizin an der Eberhard-Karls-  
Universität, Tübingen

1998 Naturwissenschaftliche Vorprüfung Zahnmedizin

2000 Physikum Zahnmedizin, Physikum Medizin

2001 1. Staatsexamen Medizin

2003 Staatsexamen Zahnmedizin, 2. Staatsexamen  
Medizin, Promotion Zahnmedizin

16.11. 2004 3. Staatsexamen Medizin