DISCUSSION
CHAPTER 5
DISCUSSION

5.1. Introduction
The quality of health care services has gained a major position in social issues. As a result the health care providers are under pressure to demonstrate the added value of the services they provide (Thomas and Stewart 2005). Practice of assessment and evaluation in physiotherapy has been emphasised not only for the purpose of quality service, but also for audit and research advances (Stavem and Arnesen 2005; Kreibich et al 1996). So it has become essential for the physiotherapist to assess the effectiveness of the treatment using an outcome measure which is valid and reliable. Lack of objective measurements may lead to improper decision making, hindering the patient’s improvement. Besides improving the quality of health care services, reliable outcome measures enhance the quality of trials in which it is used (John 2004).

A reliable outcome measure must produce consistent results after repeated measurements by examiners during different periods of time on the same patients (Thomas and Stewart 2005; John 2004). The outcome measure must be used with steps taken to reduce possible sources of variability and bias in scoring by selecting an appropriate tool, standardising the procedures and repeated practice by the examiners (Wright and Feinstein 1992). A reliable outcome measure improves the examiners confidence in deriving conclusions based on results produced by them (Wright and Feinstein 1992). The aim of this study was to establish the intra and inter-rater reliability of two examiners using the KSKS. A correlation design of test-retest
reliability was used in this study to assess the correlation between the measurements by Examiner A and Examiner B. Correlation designs are used to compare the similarities between different sets of data (Hicks 1995). The common method of test-retest reliability was implemented by administering the KSKS at two different times with an average time interval of 60 minutes between them. The ICC (John 2004) was used to establish the reliability of measurements obtained during different times by the examiners.

The time interval in this study was not so short so that the memory of the previous test biased the performance of the examiner (Thomas and Stewart 2005) and not too long so that there were no changes in the attributes which were measured (Finch 2002; Campbell et al 1999).

5.2. Intra rater reliability

5.2.1. Intra rater reliability of the Knee Society Knee Score
The results of this study revealed a high correlation of the KSKS when tested with an average time interval of one hour between the two tests. Examiner A showed excellent intra-rater reliability (h=0.95) and Examiner B showed good intra-rater reliability (h = 0.71). In a similar study (Liow et al 2000) on reliability of knee score was administered by six examiners with varying experience on 29 subjects. The study showed considerable variations in intra-rater reliability which was attributed to low experience of the examiners and lack of training in administering the tool. They also found examiners with more than three years experience showed higher intra-rater
reliability relatively. In our study both the examiners had more than four years experience and were trained in the assessment tool prior to the study. They therefore showed a high correlation between their total scores.

**5.2.2. Intra rater reliability of the individual items in the Knee Society Knee Score**

The measurements obtained in the study showed moderate to excellent correlation between the individual items of the KSKS.

**5.2.2.1 Knee Range of Motion (ROM)**

When analysing the Knee ROM component of the KSKS we found that Examiner A and Examiner B showed excellent intra-rater reliability (h=0.96 and 0.94) respectively. Smith and Walker, in 1983 reported high intraobserver reliability (r = 0.90). Reliability of the examiners in this study is higher than that reported by Smith and Walker. They also used a universal goniometer which is accepted to be reliable (Gajdosik and Bohannon 1987).

**5.2.2.2 Stability**

When analysing the stability components of the KSKS Examiner A and Examiner B showed good intra-rater reliability in AP stability and ML stability (refer Table 4.3). A previous study (Liow et al 2000) demonstrated moderate correlation in an experienced examiner with a Kappa value = 0.50. In this study both the examiners showed good correlation between their scores. This may be attributed to the fact that
both the examiners went through a practice session in which they adopted common techniques to reduce variability between their measurements.

5.2.2.3 Flexion Contracture

When analysing the flexion contracture component of the KSKS, Examiner A showed excellent intra-rater reliability (h=0.95), while Examiner B showed good intra-rater reliability (h=0.89). This is significantly higher than in a previous study (Liow et al 2000) which had a Kappa value=0.52 for the experienced observer. This may be due to the precise marking of anatomical landmarks before measuring with the goniometer by both the examiners.

5.2.2.4 Extension Lag

When analysing the extension lag component of the KSKS, Examiner A showed moderate intra-rater reliability (h=0.65), while Examiner B showed good intra-rater reliability (h=0.87).

In the individual components of the KSKS there is a high intra-rater reliability between them. It could be argued that this reliability was biased by the memory of previous measurement as the time interval (one hour) between the two tests was shorter than that of the time interval (two hours) of the study conducted by (Liow et al 2000). But the chance was minimised by recruiting an independent observer to note down the measurements from the tests with the intention that if the examiner was not actually writing down the measurement it would less easily be committed to memory. It is of interest that knee ROM and flexion contracture showed very little variation. This may
be due to the fact that both are controlled passively by the examiners and the measurements are taken from the goniometer reading which is more reliable than a clinical measure of stability.

5.3. Inter rater reliability

5.3.1. Inter rater reliability of the Knee Score

The results of this study revealed a moderate inter-rater reliability between the examiners during test 1 (h=0.67) and test 2 (h=0.66). Ryd et al (1997) reported low inter-rater reliability with a standard deviation of 26 for the knee score which is larger than that reported in this study (SD = 16). The examiners in this study standardised the measurement procedures through repeated practice, training and discussion prior to the study. This is supported by Liow et al (2000) as they reported high inter-rater reliability with trained individuals on the test procedures.

5.3.2. Inter rater reliability of individual items in the Knee Society Knee Score

The measurements showed poor to good correlation between the individual items of the Knee score.

5.3.2.1 Knee Range of Motion (ROM)

When analysing the knee ROM component of the KSKS good inter-rater reliability was found between the examiners during test 1 and test 2 (h=0.85 and 0.82). Gajdosik and Bohannon (1987) in their study reported high inter-rater reliability with r=0.70. Our study reported much higher reliability between the examiners during both
the tests. As discussed under section 5.2.2.3 the high reliability of the examiners in using the goniometer may be attributed to the precise marking of the anatomical landmarks which was not found in the study by Gajdosik and Bohannon (1987).

5.3.2.2 Stability

In the component of stability (AP stability and ML stability) of the KSKS poor inter-rater reliability was found between the examiners during test 1 and test 2 (refer Table 4.5). The scores showed comparatively low inter-rater reliability compared to that reported between experienced staff by Liow et al (2000) for ML stability (Kappa value=0.36). This may be due to the difference in the forces applied by the examiners to perceive the amount of movement at the knee joint.

5.3.2.3 Flexion Contracture

When analysing the flexion contracture component of the KSKS, moderate inter-rater reliability was found between the examiners during test 1 and test 2 (h=0.58 and 0.64). This is relatively higher than that reported between experienced staffs by Liow et al 2000 for the same component (Kappa value=0.19).

5.3.2.4 Extension Lag

The results of the extension lag component of the KSKS show moderate inter-rater reliability between the examiners during test 1 (h=0.54) and good inter-rater reliability during test 2 (h=0.76). The increased reliability between the examiners may be attributed to the fact that the patients would have become more aware of knee
extension during administration of the test for the second time. It may be also due to repeated reinforcements from the examiners to extend the knee.

The individual components of the KSKS show a low inter-rater reliability between them. The stability component showed the least inter-rater reliability, which can be due to the subjective measure of self perceived movement by examiners which tend to vary with individuals. Unlike this, the measurements from a goniometer are more objective and showed highest inter-rater reliability among the items in the KSKS.

**5.4 Summary**

In conclusion the results of this study showed good intra-rater reliability and a moderate inter-rater reliability for the KSKS that is to be used in a clinical trial. Based on the results of this study the researcher feels that the KSKS can be used reliably in a clinical trial.

Further research should be concentrated on developing a tool which combines objective measures that demonstrate high reliability or modifying the measurement tools so that certain components (stability) of score can become more objective.