Table 8-1.
Intrarater Reliability of Humeroulnar Flexion Range of Motion

	SUBJECT 1			SUBJECT 2			SUBJECT 3			SUBJECT 4		
Trial	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Rater A	145	142	142	138	138	137	140	141	142	132	132	132

Table 8-2.
Interrater Reliability of the Balance Error Scoring System

	BESS SCORE SUBJECT 1 (TOTAL NO. OF ERRORS)	BESS SCORE SUBJECT 2 (TOTAL NO. OF ERRORS)	BESS SCORE SUBJECT 3 (TOTAL NO. OF ERRORS)
Rater A	12	20	18
Rater B	14	23	19
Abbreviation: BESS, Balance Error Scoring System.			

consistently higher scores than rater A each time, but the discrepancies between the raters are not drastically different. To determine if the raters have good reliability, an intraclass correlation coefficient (ICC) should be used.

Statistical Application for Rater Reliability

An ICC is most commonly used to determine if the reliability of one rater or a group of raters is consistent. The ICC is a reliability coefficient statistical measure that is sensitive to changes between raters and/or between trials. There are 3 common ICC models that are used to assess reliability data (Table 8-3). This table explains how raters are selected and how data are assessed for each ICC model. Additionally, the table provides the mathematical formula for calculating the ICC value for each model. When discussing the ICC models, it is important to understand that the values in the parentheses indicate the model and form, respectively. An ICC (2,1) indicates that it is model 2 and form 1. The **model** (first number) differs in accordance to how the raters are chosen within a study and how they are assigned to the subjects. The **form** (second number) is determined if you are using an average of measures (referred to as k) or a single measure (referred to as 1).

The practice of using an ICC (1,1) or (1,k) model is rarely performed in reliability studies because each subject is being assessed by a different rater. It would be extremely difficult to assess the reliability if all subjects were being assessed by a different rater. Intraclass correlation coefficient models (2,1) and (2,k) are chosen when the subjects and raters are considered to be randomly selected from the larger population, thus allowing the results to be generalized outside the study. For example, if you wanted to assess the ability of athletic training students to accurately assess dorsiflexion range of motion, you would randomly select students from all over the country. The students are selected from the larger population, so an ICC model 2 would be best because those students are a representative sample of all athletic training students. Intraclass correlation