ESTABLISHING THE RELIABILITY OF TOTAL BODY CENTER OF GRAVITY VALUES

Sarah L. Smith
Health, Physical Education and Sport Studies
Rutgers University
New Brunswick, New Jersey 08903

Marilyn A. Looney
Physical Education and Dance
University of Iowa
Iowa City, Iowa 52242

At the CIC Big Ten Symposium on Biomechanics in October, 1980, two of the presentations focused on measurement and statistical considerations in biomechanical research. In both instances, statements were made with respect to the concept of the reliability of biomechanical data. Morris (1980) referenced the observations of Kroll who had noted the "poor assessment of reliability and objectivity" (p. 233) in biomechanics research. Disch and Hudson (1980) also noted their concern by stating that "the reliability questions important to the biomechanics researcher are ones of stability and objectivity" but "unfortunately, this important measurement phase of biomechanical research is often overlooked." (p. 146)

Researchers in sport biomechanics often determine center of gravity (COG) values which are calculated using the segmental method. These COG values subsequently provide basic information (e.g., displacement, velocity, and acceleration) about a specific movement or skill being analyzed. Although the concept of using calculated COG values for describing human performance is routinely accepted in biomechanical research studies, there is little evidence in the biomechanics literature to counter the criticisms cited above regarding the reliability of the COG data that has been reported.

Although error sources within the data collection phase of cinematographical studies have been identified, no attempt has been made to quantify the magnitude of these error sources or to suggest an appropriate measurement procedure for ensuring reliable COG data. Additionally, the previous studies have been unidimensional in scope - only intraplotter or interplotter error has been analyzed. Since several sources of error are associated with obtaining the digitized X- and Y-coordinate values, a flexible technique should be used to estimate the reliability of these measurements. The application of generalizability theory, formulated by Cronbach et al. (1972), provides such a technique. It is the purpose of this presentation to examine previous methods used to establish the reliability of center of gravity values in biomechanical research and to propose an alternate method for obtaining reliability estimates.

Intraclass correlations coefficients were reported in the investigations of Barlow (1973), Ward (1973), Davis (1973), and French (1981). These correlations were calculated separately for the repeated measures of the X- and Y-coordinate values of the segmental endpoints or anatomical landmarks digitized in each study. Both Barlow and Ward digitized two trials whereas Davis and French digitized three trials of the segmental endpoints. In all instances, the reported intraclass correlation coefficients ranged from .91 to .99 for both the X- and Y-coordinates. These intraclass correlation values, however, reflect only the stability or consistency aspect of reliability.

Davis (1973) was unable to compute an intraclass correlation coefficient that would represent the objectivity aspect of reliability. An analysis of variance of his data indicated that agreement among plotters in estimating both the X- and Y-coordinates for the location of total body COG did not exist.

Before illustrating how generalizability theory provides a way of estimating the reliability of center of gravity values, it is necessary to review some of the basic aspects of this theory. Some comparisons between generalizability theory and classical test theory will also be noted.

Generalizability theory recognizes that a person's score is representative of the conditions of measurement under which that score value was obtained. This score reflects performance within a specified universe or condition(s) of measurement. From the information obtained in a generalizability or G study, it is possible to generalize in a decision or D study to a different universe about a particular measurement procedure.

In any measurement procedure, generalizability or G theory decomposes a person's score into a universe score component and one or more error components with respect to the grand mean. The universe score is an average score value over all the conditions of the measurement procedure. Each condition specified by the investigator(s) is viewed as a facet of measurement which represents a possible source of error. The objects of measurement, which are usually subjects and are referred to as persons, are not viewed as a facet and therefore, are not a source of error.

Variances can be estimated for the objects of measurement, each facet, and all interactions of these factors; and when summed, these variances represent the estimated variance of the observed scores. These variance components are the focus of a G study.

The magnitude of the sources of variance identified in a measurement procedure can be assessed by examining their estimated variance components which are obtained through the appropriate analysis of variance design. "The relative magnitudes of these components provide information about particular sources of error influencing a measurement." (Shavelson and Webb, 1982, p. 133) Additionally, a percentage value that each source of variance contributes to the total variance of the measurement can be determined.

Brennan (1983) states that "these sources of variance are called G study variance components" (p. 5), and they are used to determine both universe score and error variances. Different types of error variances are recognized in generalizability theory, but two that are most often discussed are relative error variance [$\sigma^2(\delta)$] and absolute error variance [$\sigma^2(\Delta)$]. The distinction between these two types of error variance is that relative error variance is used when norm-referenced or comparative interpretations are of central importance. When criterion- or domain-referenced interpretations of scores is of interest, absolute error variance is required.

By definition, reliability is equal to true score variance divided by the sum of true score variance and error score variance. In generalizability theory, the ratio between universe score variance and expected observed score variance determines a reliability-like coefficient known as a generalizability coefficient. Since the expected observed score variance includes error variance, the appropriated type of error variance must be used. If relative error variance is used, the resulting coefficients are known as G-coefficients; if absolute error variance is used, then the calculated ratios are identified as indices of dependability.

Several similarities between generalizability theory and classical test theory are now apparent. Universe score variance is analogous to true score variance, and a generalizability coefficient is technically an intraclass correlation coefficient (Brennan, 1983). Additionally, standard deviations of the absolute error variances function similarily to the standard errors of measurement. Confidence intervals about a universe score or an individual score can thus be established.

Likewise, certain differences exist between G theory and classical test theory. Not only can multiple sources of error be differentiated in generalizability theory but also the magnitude of these error sources can be assessed. In classical test theory, the focus is on the reliability coefficient but this is not true in generalizability theory. Here, the variance components are of central importance. Also, classical test theory recognizes only one true score for a particular application whereas G theory provides as many universe scores as there are defined universes of generalization.

Additionally, there are several other advantages to using generalizability theory for establishing the reliability of the measurement procedure. G theory is not restricted to norm-referenced interpretations. By selecting to use absolute error variance in the denominator of the generalizability coefficient, the reliability for the criterion-referenced interpretation of scores is possible. Also, since several reliability-like coefficients are determined in G studies, information is provided which permits the investigator(s) to modify and improve the design of the measurement procedure.

To assist with the interpretation of G-study results, Brennan and Kane (1977) have suggested using signal-noise ratios.

By comparing universe score variance to <u>either</u> absolute or relative error variance, a signal-noise ratio is formed. This ratio provides an index of the relative precision of the measurement procedure for domain-referenced or comparative interpretation of scores. If the signal is large compared with the noise, a large ratio results and is indicative of the adequacy of the measurement procedure.

One important aspect of G theory relates to the simultaneous treatment of multiple universe scores. This feature would permit the biomechanics researcher to examine the reliability of COG X- and Y-coordinates simultaneously. As Brennan (1983) has noted, "the multivariate feature of generalizability theory is one of its unique characteristics." (p. 113)

As an example, generalizability theory has been applied to determine the reliability of COG values obtained by the segmental method under two different measurement conditions. The two facets in the G study were digitizers or plotters and sequences for digitizing the segmental endpoints.

Twenty-eight college-aged students (males = 14; females = 14) were filmed by a LOCAM camera at 100 fps. Mean height and weight values for the males were 178.79 cm and 77.50 kg. The corresponding values for the females were 163.83 cm and 57.74 kg. All subjects were attired in shorts, short sleeve shirts, and athletic shoes and filmed while performing the basic locomotor skill of walking.

Film analysis was conducted on each subject using six frames of film depicting a one-stride walking cycle consisting of right heel strike, right foot flat, left toe off, left heel strike, left foot flat, and right toe off. All film frames analyzed were marked to ensure that identical frames were digitized by the two plotters who used two different digitizing sequences on alternate days. The same digitizing unit was used in the data collection process.

Nineteen segmental endpoints and one reference point were digitized in a specified order for each of the six film frames in Sequence 1. In Sequence 2, one anatomical landmark was digitized throughout the entire six frames of the stride cycle; then a 2nd, 3rd, etc. anatomical landmark was digitized throughout the entire six frames of the stride cycle. This sequence had been selected because of the suggestion that anatomical landmarks, particularly those hidden by other body parts, could be located more accurately if they were tracked throughout the entire movement sequence being analyzed. There was nearly a four-fold increase in the time required for the digitizing process using Sequence 2 as compared with Sequence 1.

Four COG values were determined for each subject in each of the six positions of the stride. Two similar FORTRAN computer programs used the same body segment parameters for calculating the COG values. Because of the two different digitizing sequences, the computer programs varied only with respect to the order of reading the segmental endpoint data. Also, an identical reference point was used in both programs and provided a common origin with respect to the four sets of COG values. The average values of the four X- and Y-coordinates for each subject constituted the X and Y universe scores for that subject.

The X- and Y-coordinates for these COG values were analyzed separately by the BMD8V computer program using a fully crossed 3-way ANOVA design. Digitizer (plotter) and sequence facets were considered as being random. Estimated variance components and percentages of total variance for the seven sources of variation were computed. These estimated variance components provided the means for examining the reliability of the three specified universes of generalization. In Table 1, the estimated variance components for Frame 1 are presented.

TABLE 1: ESTIMATED VARIANCE COMPONENTS AND PERCENTAGES OF TOTAL VARIANCE BY X- AND Y-COORDINATES FOR FRAME 1

Source of Variation	X	% of Total	Y	% of Total
Persons	1.7226	78.5	0.1588	80.4
Digitizers	--	--	--	--
Sequences	0.0261	1.2	--	--
P x D	0.0403	1.8	0.0000[a]	0.0
P x S	0.0381	1.7	0.0000[a]	0.0
D x S	0.0788	3.6	0.0063	3.2
P x D x S x e	0.2885	13.1	0.0323	16.4

[a]Negative variances were replaced with zeros (Brennan, 1984).

In this particular G study, the initial partitioning of total variance resulted in several interaction terms having larger variances than their main effects. This indicated the linear model was too elaborate. A different linear model was then defined with fewer terms. Also, certain interaction facets had negative variances and were replaced by zeros according to the procedure recommended by Brennan (1984).

As can be observed, the major contributor to score variance was the variation among persons. For the X- and Y-coordinates, the percentage of total variance contributed by persons was 78.5 and 80.4%, respectively. Residual error (P x D x S and e) was the second largest contributor to score variance.

The estimated variance components were used to compute both universe score and absolute error variances. Absolute error variance was selected because a domain-referenced interpretation of the scores was desired. Indices of dependability were then determined and those for Frame 1 are presented in Table 2.

TABLE 2: INDICES OF DEPENDABILITY FOR THE THREE UNIVERSE CONDITIONS

Frame	Coordinate	Universe Score Conditions[a]		
		Condition 1 (D=2,S=2)	Condition 2 (D=2,S=1)	Condition 3 (D=1,S=1)
1	X	.923	.865	.785
	Y	.943	.892	.804

[a]D = Number of Digitizers; S = Number of Sequences.

These indices of dependability reflect the reliability of the measurement procedure under specified universes of generalization. Generalizing over digitizers and sequences, the index of dependability for the COG X- and Y-coordinates is .932 and .943, respectively. In Condition 2, dependability decreased to .865 for the COG X-coordinate and to .892 for the COG Y value. However, if one generalizes over the universe specified in Condition 3, the index of dependability decreases to .785 and .804 for the COG X- and Y-coordinates, respectively. It should be noted that Condition 3 reflects the usual digitizing process.

Standard absolute errors for the three universe score conditions are shown in Table 3. The imprecision of the measurement procedures is reflected by the increase in the magnitude of these values. Particular attention should be given to Condition 3 which represents the usual digitizing scheme of one plotter using one digitizing sequence. With a 68% confidence interval, the individual's universe COG value could be within ±20.93 cm in the horizontal plane and ± 5.99 cm in the vertical plane.

TABLE 3: STANDARD ABSOLUTE ERRORS FOR THE THREE UNIVERSE SCORE CONDITIONS

Frame	Coordinate	Universe Score Conditions[a]		
		Condition 1 (D=2,S=2)	Condition 2 (D=2,S=1)	Condition 3 (D=1,S=1)
1	X	11.57	15.78	20.93
	Y	3.00	4.24	5.99

Note. The 68% confidence interval for a universe score is $\bar{X}_p \pm \hat{\sigma}_\Delta$. The unit of measure is centimeters.

[a]D = Number of Digitizers; S = Number of Sequences.

Signal/noise ratios also give evidence of the precision of the measure procedure (Brennan and Kane, 1977). From the data in Table 4, the ratio for the strength of the signal in comparison with the noise shows a decrea~e from 11.96 to 3.65 for the COG X-coordinate. The corresponding values for the COG Y-coordinate ranged from 16.46 to 4.11. It is quite obvious that as the number of measurements was reduced to one at each segmental endpoint, the signal was considerably reduced.

TABLE 4: SIGNAL/NOISE RATIOS FOR THE THREE UNIVERSE SCORE CONDITIONS

		Universe Score Conditions[a]		
Frame	Coordinate	Condition 1 (D=2,S=2)	Condition 2 (D=2,S=1)	Condition 3 (D=1,S=1)
1	X	11.96	6.43	3.65
	Y	16.46	8.22	4.11

[a]D = Number of Digitizers; S = Number of Sequences.

The above application of generalizability theory has illustrated a method for determining the reliability of COG values calculated by the segmental method. As Brennan (1983) has stated, the theory "offers an extensive conceptual framework and a powerful set of statistical procedures for addressing numerous measurement sources" and it "can be viewed both as an extension of classical test theory and as an application of certain analysis of variance procedures to measurement models involving multiple sources of error." (p. 1)

Since the methodology is available, how can biomechanists fail to respond to the criticisms about the reliability of the data presented in many of their studies? Without acceptable levels of reliability, both in terms of stability and objectivity, the validity of the data is also in question. As Disch and Hudson (1980) noted, "if the findings and conclusions of biomechanical research studies are to be added to the realm of scientific literature then basic measurement questions like these of reliability and validity must be considered." (p. 201)

References

Barlow, D. A. (1973). Kinematic and kinetic factors involved in pole vaulting. Unpublished doctoral dissertation. Indiana University.

Brennan, R. J. (1983). Elements of generalizability theory. Iowa City, IA: American College of Testing Program.

Brennan, R. J. (1984, April). Some statistical issues in generalizability theory. Paper presented at the SIG/Educational Statisticians symposium at the meeting of the American Educational Research Association, New Orleans.

Brennan, R. J., & Kane, M. T. (1977). Signal/noise ratios for domain-reference tests. Psychometrika, 42, 609-625.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.

Davis, M. W. (1973). Quality of data collected by the segmental analysis technique. Unpublished doctoral dissertation, Indiana University. Microform Publications No. 612.76, College of Health, Physical Education and Recreation, University of Oregon.

Disch, J. G., & Hudson, J. L. (1980). Measurement aspects of biomechanical analysis. In J. M. Cooper & B. Haven (Eds.), Biomechanics Symposium Proceedings (pp. 191-201). Indiana State Board of Health.

French, E. K. (1981). Validity and reliability of kinematic data from two dimensional cinematography. Unpublished doctoral dissertation, Indiana University. Microform Publications No. 778.53, College of Health, Physical Education and Recreation, University of Oregon.

Morris, H. H. (1980). Statistics and biomechanics: selected considerations. In J. M. Cooper & B. Haven (Eds.), Biomechanics Symposium Proceedings (pp. 216-225). Indiana State Board of Health.

Shavelson, R. J. & Webb, N. M. (1981). Generalizability theory: 1973-1980. British Journal of Mathematical and Statistical Psychology, 34, 133-166.

Ward, P. E. (1973). An analysis of kinetic and kinematic factors of the standup and the preferred crouch starting techniques with respect to sprint performance. Unpublished doctoral disseration, Indiana University. Microform Publications No. 796.426, College of Health, Physical Education and Recreation, University of Oregon.