On the Problem of Two-Dimensional Error Scores: Measures and Analyses of Accuracy, Bias, and Consistency

Gregory R. Hancock  
Department of Educational Foundations, Leadership, and Technology  
Auburn University

Michael S. Butler  
Division of Health, Physical Education, and Recreation  
Emporia State University

Mark G. Fischman  
Department of Health and Human Performance  
Auburn University

ABSTRACT. Describing and analyzing error for one-dimensional performance tasks is fairly straightforward, but suggestions for describing and analyzing error for two-dimensional performance tasks (e.g., marksmanship) are quite problematic. Specifically, imposing an arbitrary axis onto the two-dimensional work space, along which traditional one-dimensional measures can be computed and analyzed, yields measures of accuracy, bias, and consistency that are entirely dependent upon the choice of axis. The present contribution offers new measures and methods for describing and analyzing data from two-dimensional performances. Unlike the results from previous suggestions, the approaches described herein yield results that are completely independent of the axes used to quantify the individual two-dimensional trials. These new approaches are strongly related to well-established methods for describing and analyzing error for one-dimensional tasks.

Key Words: error scores, measurement, two-dimensional performance

Motor behavior research cannot be accomplished without the quantification of subjects' performance on some task. This quantification has been achieved for certain kinds of tasks through the use of error scores designed to reflect various aspects of performance. Traditionally, two categories of error scores have been employed to describe performance: measures of accuracy and measures of consistency. Of the different measures of accuracy used, the most common are constant error (CE) and absolute error (AE). CE is the signed error for a single trial, but usually is reported as the mean CE averaged over a block of trials for 1 subject. AE, on the other hand, is the unsigned error for a given trial and, therefore, ignores direction of bias in a trial. Like CE, AE is usually reported as the mean AE over a block of trials for 1 subject (unless explicitly stated otherwise, within this article the terms CE and AE refer to the average across trials for a single subject). The fundamental difference between AE and CE, however, is that CE accounts for the sign (i.e., direction with respect to the desired target) of each trial and thereby represents the magnitude and direction of the bias of a subject's positionally typical trial.

When researchers wish to aggregate measures of accuracy or bias across subjects, the properties of the different measures of accuracy must be considered. An aggregation of AE measures across subjects is quite straightforward: Subjects' AE measures simply are averaged. For CE, however, such an average may be considered misleading because the positive and negative biases of different subjects cancel each other out; thus, an average CE across subjects tends to converge upon zero. Because of this, many investigators have turned to reporting absolute constant error (ACE) as the preferred descriptor of across-subjects accuracy (Schutz, 1977), where ACE is the mean of all subjects' unsigned CE measures.

Even if accuracy and bias measures are computed and interpreted correctly, they are not by themselves adequate for a complete description of performance. Consis-

Correspondence address: Gregory R. Hancock, Department of Measurement, Statistics, and Evaluation, University of Maryland, Benjamin Building, Room 1230, College Park, MD 20742-1113.
tency is the other dimension of performance that must be described, the traditional measure of which is variable error (VE). VE can be computed for a single subject across a block of trials as the standard deviation of that subject’s CE scores for each trial (Magill, 1993), or equivalently, the standard deviation of the subject’s raw scores for each trial (Schmidt, 1988). When an aggregate measure of consistency is desired across subjects, subjects’ VE measures simply are averaged. This serves as a reasonable indicator of the average extent to which subjects deviated from their own positionally typical trial.

Although there has been some debate about the appropriate use and interpretation of the various error measures (see Schutz, 1977; Schutz & Roy, 1973), motor learning research has proceeded relatively unhindered by the lack of consensus on these measures. For one-dimensional (1-D) tasks such as linear positioning, or tasks in which time is the variable of interest, such as coincidence timing, the application of the various error measures is fairly straightforward. When a researcher is faced with the quantification of error for two-dimensional (2-D) tasks such as marksmanship, dart throwing, or archery, however, extending the 1-D error measures to describe performance in two dimensions is not so straightforward.

Suggestions for dealing with 2-D error scores have been put forth; however, certain problems have yet to be resolved. For example, Magill (1993) described a situation in which an instructor wanted to provide feedback to students in an archery class. Magill suggested that an arbitrary division (i.e., vertical or horizontal) of the 2-D work space (the target) would yield a usable frame of reference for the calculation of traditional error scores. However, because a subject’s tendencies in the vertical dimension are generally independent of those in the horizontal dimension, the choice of axis (vertical, horizontal, or otherwise) can have a great impact on the measures of accuracy, bias, and consistency derived from that arbitrary division. Hence, the dependence of such error scores on the choice of axes, using Magill’s suggestion, renders their informational value tenuous at best. This issue was described by Reeve, Fischman, Christina, and Cauraugh (1994).

As Schutz stated, the problem of 2-D error scores “gets very complex” (Proceedings of the Colorado Measurement Symposium, 1977, p. 107). The potential for complexity, however, should not lead us to infer that 2-D performances are unimportant, nor should it preclude us from their investigation. Such performances may be, in many cases, at least as ecologically valid as the 1-D tasks that have been the focus of many motor learning studies. Our purpose in this article is to present appropriate error measures for quantifying accuracy, bias, and consistency for 2-D performance tasks. To facilitate the development of some of these measures, we present an archery example in Figure 1. Four subjects, referred to as a, b, c, and d, have each taken four shots at the target. The center of the target (the bull’s-eye) is the goal of this task, and the solid black circles represent the shots of each subject. These four sets of four shots have been chosen for illustrative purposes and are discussed in the context of accuracy, bias, and consistency. Two-dimensional analogs to existing 1-D error measures are presented, as are methods of analysis using these error measures.

The reader will notice that specific axes appear adjacent to the target in Figure 1. These have been chosen as a frame of reference for describing shots in the 2-D space, and as is discussed later, this choice of axes has no impact on the measures and methods presented herein. Furthermore, it is important to note that the origin of these axes (0,0) has been chosen to correspond to the center of the target, or bull’s-eye. Setting the center of the target at other coordinates would in no way compromise the measures and analyses presented in this article, but placement of axes at (0,0) is descriptively attractive and makes many of the equations we present somewhat simpler, because coordinates for the natural origin drop out. In the interest of simplicity, then, throughout this article it is assumed that the researcher has selected a pair of perpendicular axes with a desirable measurement scale, and that the origin of those axes is at the center of the target.

*Figure 1.* Shots for Subjects a, b, c, and d on a target with imposed axes, and shot coordinates. Numbers in parentheses are the x and y coordinates of the individual shots.
Hence, because the measures and methods discussed herein are invariant across axial orientations, they overcome the limitations of previously proposed solutions to the problem of 2-D error scores.

Accuracy and Bias in 2-D

Accuracy

For a Single Shot

In the field, the accuracy of a single shot is roughly quantified by the concentric circles on the target itself, providing a rough integer measure of the shot’s distance from the center of the target. Using a ruler or other linear scale, this distance could certainly be measured more precisely, and such a precise quantification may be described as a measure of radial error (RE). RE serves as the 2-D analog of the 1-D AE measure for a single trial—neither associates any direction with the error, only the error’s magnitude is captured. RE could be derived for a single shot either by direct measurement or by using the shot’s coordinates \((x, y)\) on a pair of imposed \(X\) and \(Y\) axes. Given the aforementioned assumption that the bull’s-eye is at the origin \((0,0)\), it follows from the Pythagorean theorem that

\[
RE = (x^2 + y^2)^{1/2}.
\]

The reader should note that although any other orientation of axes will change the coordinates of the shot in question, given the same sized units on the imposed axes, an identical measure of RE will result.

For Summarizing Multiple Shots From a Single Subject

Once \(X\) and \(Y\) axes have been imposed by the researcher, coordinates for each shot in a block of shots may be determined. From this point the researcher’s aggregation options depend upon the question being asked. If the researcher simply wants to know how far from the bull’s-eye the subject’s shots were on average, the mean radial error \((MRE)\) may be calculated as follows. For \(m\) shots,

\[
MRE = \overline{RE} = (1/m) \sum_{i=1}^{m} (RE).
\]

The \(MRE\) measure is the 2-D analog to the \(AE\) measure across trials in 1-D, and as such it has the same limitations. Specifically, it completely ignores the position of the shots, as is evidenced by Subject a and Subject b in Figure 1. Although both have the same \(MRE\), Subject a is shooting around the bull’s-eye, whereas Subject b is consistently biased in one direction. In 1-D, this problem illustrates the difference between \(AE\) and \(CE\). \(CE\) represents the magnitude and direction of bias for a positionally typical trial and is appropriate for answering questions about where, on average, a subject’s trials tend to be with respect to the center of the target. This is a subtle, but critical, difference from \(AE\). In 1-D, then, the measure of \(CE\) accommodates differences in direction of bias.

Bias

For Summarizing Multiple Shots From a Single Subject

For a single shot, information about the magnitude and direction of bias is contained in the position of the shot relative to the bull’s-eye. However, researchers’ interest in bias usually focuses on a block of shots, rather than at the single-shot level. To derive a measure of bias in 2-D, then, one must first establish the representation of a positionally typical shot across a block of shots. Given the coordinate axes imposed upon our target, this is accomplished in a straightforward manner by using what is known in multivariate statistics as a centroid (see, for example, Tatsuoka, 1988). The centroid represents a point whose coordinates are given by the average \(x\) value and average \(y\) value of the shots being aggregated. Specifically, the coordinates of a subject’s centroid for \(m\) shots is given by

\[
(x_c, y_c) = \left( \frac{1}{m} \sum_{i=1}^{m} x_i, \frac{1}{m} \sum_{i=1}^{m} y_i \right).
\]

The centroid is not, by itself, a measure of bias in two dimensions. But, as a positionally typical trial, the centroid contains information about the subject’s overall magnitude and direction of bias, and from this point a measure of magnitude of bias may be derived. Quite simply, given that the bull’s-eye is at the origin \((0,0)\), the subject-centroid radial error \((SRE)\) may be computed as

\[
SRE = (x_c^2 + y_c^2)^{1/2}.
\]

This measure represents the radial distance of the centroid from the bull’s-eye and is a measure of the magnitude of bias over multiple shots from a single subject. The centroids for Subjects a, b, c, and d, and their respective measures of \(SRE\), are given in Figure 2.

From Figure 2 one notices that Subject a has a centroid directly on top of the bull’s-eye, because there is no systematic bias among the four shots. The \(SRE\) for Subject a is thus zero. Subjects b and c, however, whose shots are in different locations on the target, have measures of \(SRE\) identical to each other. This is because their centroids are equidistant from the target’s center, and the \(SRE\) measure contains no directional information. \(SRE\) may thus be considered the 2-D analog of \(ACE\), conveying information only about the magnitude of bias.

In 1-D, the communication of direction of bias is done through the sign attached to the \(CE\) measure. In 2-D, the addition of a sign to the \(SRE\) measure would be virtually without meaning; and, even if one could attach some system of signs to the \(SRE\) measure that would communicate differences in direction of bias, this system would no doubt be dependent upon the arbitrarily imposed \(X\) and \(Y\) axes. The use of an angular deviation measure, as in
polar coordinates, complements the MRE measure of magnitude of bias and seems to offer an attractive approach to communicate direction of bias. This solution, however, has two serious limitations. First, in aggregating angular deviations across shots (or across subjects) to represent a typical direction of bias, misleading results arise with angles to both sides of the axis of angular reference (0°). For example, the numerical average of 10° and 350° (two angles near 0°) is 180°, on the opposite side of the target; likewise, using negative deviations for angles "below" the reference point would be problematic for averaging, say, 170° and −170°. Second, and equally problematic, is that all angular deviations would be completely dependent upon the choice of a reference axis. Our goal in the present article is to derive 2-D error measures that are invariant across the choice of axes. Given any system of axes with the same sized units as that used here, it is easily seen that SRE is one such invariant measure of the magnitude of bias. As for communicating direction of bias for a single subject by using the methods proposed in this article, such communication unfortunately exists only via visual inspection of the centroid's position on the target itself. However, a method of testing for significant positional differences between groups of subjects, which includes differences in direction of bias, is presented later in the article.

For Summarizing Multiple Subjects From a Single Group

Just as a subject centroid captures the positionally typical shot, a group centroid can be used to capture the positionally typical subject. That is, the group centroid of all subjects' centroids illustrates where, on average, subjects in a given group tended to shoot. Formulaically, if each subject's centroid is given by the coordinates \((x_i, y_i)\), then the coordinates of the group centroid for \(n\) subjects is simply

\[
(x_{g}, y_{g}) = (\bar{x}, \bar{y}) = \left( \frac{1}{n} \sum_{i=1}^{n} x_i, \frac{1}{n} \sum_{i=1}^{n} y_i \right)
\]  

(5)

This is, in fact, the centroid of all shots taken by subjects in a particular group, given that those subjects all took the same number of shots.

Now, assuming the bull's-eye is at the origin (0,0), the magnitude of bias of the group centroid, or group-centroid radial error (GRE), is quantified as

\[
GRE = (x_{g}^2 + y_{g}^2)^{1/2}
\]  

(6)

Referring to Figure 2, the group centroid is represented by the black square, whose coordinates are \((−0.5, 0.5)\). Applying the above equation, the GRE is computed to be 0.707. This represents the distance of the group centroid from the bull's-eye, in other words, the magnitude of bias of the group as a whole.

The reader will note that finding the GRE is analogous to averaging subjects' CE measures in 1-D. Because subjects with positive and negative biases can compensate for each other, the average of all subjects' CE measures represents a positionally typical subject. In 2-D, the positions of subjects' centroids within a group compensate for each other when determining the group centroid. Thus, the obvious limitation that comes in this type of aggregation across subjects within a group is this: If subjects' centroids are distributed around the target, the group centroid will begin to converge upon the bull's-eye.

Consistency in 2-D

In 2-D, as in 1-D, a subject's performance across multiple trials is only partially described by a measure of accuracy or a measure of magnitude of bias. In Figure 2, we notice Subject b and Subject c have identical SRE measures, though in Figure 1 we see differences in their clusterings of shots. One requires a measure of consistency in addition to a measure of magnitude of bias to describe each subject's performance more fully.

In 1-D the accepted measure of consistency is \(VE\), where \(VE\) is the standard deviation of trials from the positionally typical trial. In 2-D the same concept exists. The centroid of a block of shots represents the positionally typical trial, a mean in two dimensions, and the distance of each shot from that centroid is a deviation.
Thus, a standard deviation measure of intrasubject variability across 2-D trials, the bivariate variable error (BVE), is given by the square root of the $k$ shots' mean squared distance from their centroid $(x_j, y_j)$. Formally,

$$BVE = \sqrt{\frac{1}{k}(\sum_{i=1}^{k}(x_i - x_j)^2 + (y_i - y_j)^2)}$$

(7)

Because this measure is used only descriptively, rather than as an inferential estimator of a population variability parameter, no adjustment for lost degrees of freedom is made.

From Figures 1 and 2, it is clear that Subject a has the greatest variability of shots around the centroid; Subjects c and d have the same dispersion, and Subject b has the least. These observations are borne out through application of the preceding equation: $BVE_a = 5.09846$, $BVE_b = 0.70711$, $BVE_c = 1.41421$, and $BVE_d = 1.41421$. Readers are left to verify these calculations for themselves. And again, it should be obvious that this error measure, like all others presented, is independent of the position of the imposed axes. Given that units of the same size are used, the dispersion of shots around the centroid is not affected by changes in the placement of the coordinate system.

**Methods of Analysis Using 2-D Error Measures**

In motor behavior research, as in all research, choosing the appropriate method of data analysis is critical to properly answering the question of interest. That choice is, in fact, entirely dependent upon the question of interest. Thus, this portion of the article has been organized around various question formats with which a researcher may be faced. For each question the mechanics of the appropriate 2-D data analysis are given, as is justification for each method's appropriateness.

The reader should take care to note the differences among questions and should exercise extreme caution in choosing the appropriate analysis. Many questions appear similar, but they have fundamental differences; applied researchers are encouraged not only to make certain that one question is appropriate for their needs, but also that the other seemingly similar questions are not. Figures are presented that help illustrate the differences among different questions. Although the number of shots and subjects in these diagrams are too small for any meaningful analysis, the diagrams should help clarify conceptual differences among the questions and analyses presented.

**Questions of Accuracy and Bias**

**Do the Shots From Subjects in Different Groups Differ in Their Accuracy?**

The emphasis in this question is on the accuracy of the shots themselves. Recall that the MRE measure (see Equation 2) captures a subject's average shot distance from the bull's-eye, regardless of the shots' positions on the target. Thus, an analysis of variance on subjects' MRE measures will answer questions about differences in accuracy of the shots from subjects in different treatment groups. Considering such an analysis across Groups 1 and 2 in Figure 3, the subjects' shots in Group 1 are about as accurate (i.e., as far from the bull's-eye) as those in Group 2, and thus would not lead to any significant group differences. Though this comparability is not

---

*FIGURE 3. Groups with similar subject accuracy, different magnitudes of subject bias, similar magnitudes of group bias, and different subject consistency.*
immediately apparent upon inspection of the figure, it becomes more evident upon rotating all subjects' shots in each group to a common axis of the target (while preserving each shot's distance from the bull's-eye).

The reader should recognize that \( MRE \) describes accuracy in 2-D in much the same way as \( AE \) does for 1-D tasks, and as such the same interpretational caveats are required. There are circumstances when \( AE \) in 1-D is, in fact, the appropriate unit of analysis (Schutz, 1977). For example, as Schmidt (1975) has pointed out, the use of \( AE \) would be appropriate when the magnitude of error can be justified as the only measure of interest; a similar recommendation, then, holds for \( MRE \) in 2-D. Like \( AE \), the \( MRE \) measure is not affected by the direction of the individual trials. Therefore, two groups could have similar \( MREs \), as do Groups 1 and 2 in Figure 3, and yet differ greatly in the shots' positions on the target. Thus, the researcher must recognize that \( MRE \) does not provide a complete picture of performance and may therefore be misleading when subjects differ widely in terms of their shots' positions. If such differences are of potential concern to the researcher, the next analysis may be more appropriate.

**Do the Subjects in Different Groups Differ in Their Magnitude of Bias?**

A subject's centroid across a block of shots represents his or her positionally typical shot, the magnitude of bias of which is captured in the \( SRE \) measure (see Equation 4). An analysis of variance on subjects' \( SRE \) measures across different treatment groups, therefore, will address this question of group differences in subjects' magnitude of bias. In Figure 3, the subjects in Group 1 have centroids very near the bull's-eye. They should have a significantly lower magnitude of bias than those subjects in Group 2, whose centroids are considerably farther from the target's center.

Using the \( SRE \) measure as the unit of analysis is akin to using the \( ACE \) measure in 1-D analyses. That is, subjects' individual biases are taken into account, as seems desirable for most research purposes. Given that treatments are imposed upon subjects, where those subjects' shots tend to fall is of primary interest. Thus an analysis involving the subject centroid, which captures this tendency for each subject, is deemed appropriate. The limitation of such an analysis exists, however, when the researcher's interest is specifically in the behavior of groups of subjects as whole entities. Such an interest seems to be less common because we do not often have reason to postulate bias of groups as a whole. Nonetheless, if the researcher wishes to investigate bias of the groups, as opposed to bias of the subjects within the groups, the next questions and suggested analyses may better suffice.

**Do the Groups Differ in Their Magnitude of Bias?**

This question, unlike the two previous and more common questions, builds upon the relationship among subjects within their respective treatment groups. Recalling that a group centroid illustrates where, on average, subjects in a given group tended to shoot, the question of interest involves comparing distances of those group centroids from the center of the target. Descriptively, such a comparison is an easy task. The group with the centroid most distant from the bull's-eye, that is, the group with the largest \( GRE \) (see Equation 6), is obviously the group with the greatest magnitude of bias.

If the researcher wishes to make no generalizations beyond the specific groups of subjects and shots studied, such a descriptive approach would be adequate. Based on Figure 3, for example, because both groups have group centroids very near the center of the target, one would conclude that the two groups do not differ in bias. However, almost always, the subjects in the different treatment groups represent samples from populations, and it is those populations who are the real groups of interest. In this case, then, groups of subjects' centroids can be thought of as random samples from a bivariate population of centroids. As a result, the group centroids represent means of samples of subjects' centroids, and given that the samples contain random variability, so, too, do the group centroids. It is this sampling variability of group centroids, or more directly, the sampling variability of their magnitudes of bias, that must be incorporated into an inferential analysis involving those group centroids.

To derive a variability measure related to magnitude of bias, one must transform the coordinates of subjects' centroids within a group to be in a single dimension where only magnitude of bias relative to the bull's-eye, and not direction of bias, is preserved. As shown in Figure 4, this is first accomplished by creating an axis passing through both the bull's-eye and the group centroid \((x_g, y_g)\). The equation of this axis is given by \( Y = (y/x)X \). Then, the coordinates of the perpendicular projection of each subject's centroid onto this new axis must be determined. For each subject's centroid \((x_s, y_s)\) within a treatment group, the coordinates of the projected centroid \([x_{proj}, y_{proj}]\) are given by

\[
[x_{proj}, y_{proj}] = \left[ \frac{x_s + (y_s/y_g)x_g}{(y_s/y_g) + (y/x_g)}, y_s + (y/x_g)y_g \right] \tag{8}
\]

Verification of these geometric relationships is left to the ambitious reader.

From these coordinates of subjects' projected centroids, a measure of variability may be derived. Specifically, how subjects' projected centroids vary from the group centroid along the imposed axis is a reflection of variability in magnitude, and not direction, of group bias. Thus, an intragroup variance measure across subjects' projected centroids, the projected error variance (\(PEV\)), is given by subjects' projected centroids' mean squared distance from the group centroid \((x_g, y_g)\). Definitionally, for a single group of \( n \) subjects.
Two-Dimensional Error Scores

\[ \text{PEV} = \frac{1}{n-1} \sum_{i=1}^{n} \left( (x_{i,\text{centroid}} - x_s)^2 + (y_{i,\text{centroid}} - y_s)^2 \right) \]  

(9)

however, the following is an equivalent and computationally simpler equation:

\[ \text{PEV} = \frac{x_s^2 S_x^2 + y_s^2 S_y^2 + 2x_s y_s r_{xy} S_x S_y}{x_s^2 + y_s^2} \]  

(10)

where \( S_x^2 \) is the variance of the original \( x \) values of subjects' centroid coordinates, \( S_y^2 \) is the variance of the original \( y \) values of subjects' centroid coordinates, and \( r_{xy} \) is the Pearson correlation between the original \( x \) and \( y \) values of subjects' centroid coordinates. The variance measures should be computed with \((n - 1)\) degrees of freedom, a necessary adjustment for the subsequent inferential applications described below. Certainly, a standard deviation measure may also be derived by taking the square root of the PEV measure; however, because this will be used later in its variance form, it is presented as such. As for the equivalence of the definitional and computational forms of the equation for PEV, verification is again left to the reader.

From Figure 4 it should be noted that other groups of subject centroids could yield the same value of PEV as the group shown in the figure. Specifically, consider a second group of subject centroids, all falling on the dashed projection lines and each being, say, half way closer to the axis through the group centroid than the subject centroids appearing in the figure. Because this second group of subject centroids would yield a group centroid identical to that of the depicted group, and because the projected subject centroids would also be identical, so too would be the measure of PEV. And yet, seemingly contradictory is the fact that these two groups of subjects (the group shown in the diagram and that group hypothetically proposed here) in fact differ in their total amount of error. However, recall that our goal was to create a measure of variability reflecting subjects' contributions only to the magnitude of bias of the group as a whole, not to the total error. Although the total error variance of these two groups is indeed different, we are considering the magnitude and direction components of error variance separately. In actuality, two such groups of subjects would not differ at all in terms of variability related to magnitude of bias; they would differ only in terms of variability related to direction of bias. The use of the projected approach proposed here is a way to isolate that variability related only to the group's magnitude of bias; thus, obtaining equivalent measures of PEV for these two groups of subject centroids is completely desirable. Another approach, such as the use of real distances from subject centroids to the group centroid rather than the distances from the projected centroids to the group centroid, would be appropriate only for the simultaneous consideration of magnitude and direction of bias because such distances contain both such error components. This approach is considered in the next section of the article, where group differences in magnitude or direction of bias, or in both, are directly tested.

So, for each treatment group the researcher has computed a measure of the group's magnitude of bias, based on the group centroid; the GRE; and a measure of subjects' variability in bias around the group centroid, the PEV. From these measures, then, an analysis of variance may be conducted to determine differences among the \( k \) groups' magnitudes of bias. The appropriate \( F \) ratio is as follows:

\[ F = \frac{\frac{1}{(k - 1)} \sum_{i=1}^{k} (\text{GRE}_i - \overline{\text{GRE}})^2}{\frac{1}{(N - k)} \sum_{i=1}^{k} (n_i - 1) \text{PEV}_i} \]  

(11)

with \((k - 1)\) and \((N - k)\) degrees of freedom, where \( k \) is the number of groups, \( \text{GRE}_i \) is the group-centroid radial error of the \( i \)th group, \( N \) is the total number of subjects summed across all groups, \( n_i \) is the sample size of the \( i \)th group, and \( \text{PEV}_i \) is the variance of the \( i \)th group of projected subject centroids. A significant \( F \) value indicates that groups differ significantly in their magnitudes of bias; in inferential terms, this implies it is unlikely that the different treatment groups come from populations whose centroids have the same magnitude of bias. In Figure 3, for example, because the group centroids are roughly equidistant from their respective bull's-eyes, we would conclude it unlikely that the two groups came...
from populations with different magnitudes of group bias.

If, in fact, the researcher's interest is at the group level, as described in this section, he or she should take care to perform this analysis as described. A seemingly more appealing approach to completing this analysis would be to compute for each subject a measure of the distance from the projected centroid to the bull's-eye. Upon these distance measures from subjects in different groups an analysis of variance would appear to answer the question of interest, and would indeed be somewhat simpler. Unfortunately, such an approach would fail if subjects' projected centroids within any given treatment group fell on both sides of the bull's-eye. In this case, the mean of the distances from the bull's-eye would not be the same as the distance from the group centroid, the GRE. This, in turn, would lead to incorrect variance measures, yielding fairly uninterpretable results. Thus, if the researcher is certain an analysis of magnitude of bias at the group level is desirable, using the equations presented and completing the computations as suggested is strongly advised.

**Do the Groups Differ in Either Their Magnitude or Direction of Bias or in Both?**

The previous question was the first to use the relationship among subjects within the same treatment groups to investigate differences at the group level. The present question is a logical extension, incorporating direction of bias along with magnitude of bias. Do the treatment groups differ from the center of the target in terms of either magnitude or direction of bias or in terms of both? Answering this question requires a test that is sensitive to group differences in three scenarios: (a) where group centroids fall in a common line with the bull's-eye, but differ in their distances from that bull's-eye (i.e., differences in magnitude of bias only); (b) where group centroids are equidistant from the bull's-eye, but differ in location on the target (i.e., differences in direction of bias only); or (c) where group centroids differ both in distance from the bull's-eye and location on the target (i.e., differences in magnitude and direction of bias).

It is critical to note that in all three of the aforementioned scenarios the group centroids described differ in their position on the target. That is, group differences in magnitude or direction of bias or in both measures simultaneously are directly reflected by positional differences in one or both of the X and Y dimensions. Thus, the initial question may be rephrased as follows without any loss of information: Do the groups of subjects differ significantly on X and Y simultaneously? This question includes differences in the X dimension alone, the Y dimension alone, or in the X and Y dimensions concurrently. If such differences exist, this is directly indicative of differences in magnitude or direction or both measures of bias of the various treatment groups.

The question of group differences in a single dimension (i.e., on a single variable) is typically answered by using an analysis of variance (ANOVA). Because group differences on two dimensions (i.e., on two variables) are being considered simultaneously, the multivariate extension of ANOVA, a multivariate analysis of variance (MANOVA), is required. Specifically, treatment groups serve as the between-subjects independent variable (or variables, as in a factorial design) and the X and Y coordinates of subjects' centroids are the within-subject dependent variables. Conceptually, this is exactly a test of differences in location of group centroids, using the pooled subject-centroid variability around respective group centroids as the error term. Such an approach would provide a direct test for group differences in bias (either magnitude or direction or both), where a significant multivariate F statistic represents the presence of such differences.

Note that in Figure 5 the group centroids in Case 1 are greatly dispersed, whereas the subject centroids within each group are not. In Case 2, however, the group centroids are tightly clustered, but subject centroids within each group are quite spread out. Recalling that a MANOVA tests the ratio of between-group variability to within-group variability in multiple dimensions (as an ANOVA does in a single dimension), such a ratio would clearly be quite large in the example depicted in Case 1. Hence, we would expect such an analysis to lead to significant differences in Case 1. In Case 2, however, the presence of a small ratio is indicative of great overlap among observed bivariate distributions of subject centroids from different groups. That is, subjects belonging to one group appear almost equally likely to belong to one or more other groups. Thus, a MANOVA in Case 2 would be expected to produce no significant group differences in magnitude or direction of bias or simultaneous differences in both.

It should also be clear that the analysis proposed provides a test that is completely invariant across the researcher's choice of X- and Y-axes. In Figure 5, any other axial orientation would change the values of subject-centroids' X and Y coordinates but would in no way change the groups' positional differences, and hence would not alter the ability of the MANOVA to detect those differences.

**Questions of Consistency**

**Do the Subjects in Different Groups Differ in Their Consistency?**

In 1-D this question is equivalent to asking whether treatment groups differ in how close their subjects' trials tended to be to their own positionally typical trials. Answering the 1-D question is approached as an analysis of variance on subjects' VE measures, where VE is computed for each subject across trials. The 2-D analog follows logically. The question of interest is whether treatment groups differ in how close their subjects' shots tended to be to their own positionally typical shots (i.e.,
subject centroids). The measure capturing each subject's consistency across a block of shots was presented earlier as $BVE$ (see Equation 7), and, as in 1-D, an analysis of variance is appropriate. Specifically, conducting an analysis of variance on subjects' $BVE$ measures would directly answer the question of whether subjects in different groups differed in their 2-D performance consistency. In Figure 3, for example, subjects in Group 2 tend to be much more consistent than subjects in Group 1. Given a more realistic sample size, whose subjects' shots follow similar patterns as those shown in the figure, significant differences in subject consistency would be expected.

**Do the Groups Differ in Their Consistency?**

Unlike the previous and more common question addressing shot consistency within subjects, this deals with subject consistency within groups. Specifically, this question asks whether subjects' centroids in some groups are more spread out around their respective group centroids than subjects' centroids in other groups. Thus, the question about differences in consistency at the group level may be answered by examining distances of subject centroids from their respective group centroids. That is, for each subject centroid within a group, the subject's within-group deviation ($WD$) is computed as

$$WD = [(x_i - x_c)^2 + (y_i - y_c)^2]^{1/2}.$$  \hfill (12)

Once all subjects have $WD$ measures with respect to their own group centroid, an analysis of variance on $WD$ measures across treatment groups will answer whether groups, not subjects within those groups, differ in their consistency. In Figure 5, for example, groups in Case 1 tend to be much more consistent than groups in Case 2. Given a more realistic sample size, where subjects' centroids follow patterns similar to those shown in the figure, significant differences in group consistency would be expected.

**Summary**

The present work offers measures and analyses for describing and investigating error for 2-D tasks. Previous suggestions for dealing with this situation (e.g., Magill, 1993) have involved the imposition of an arbitrary axis along which traditional 1-D measures could be computed and analyzed. Unfortunately, such an approach yields results that are entirely dependent upon the choice of axis. That is, a vertically imposed axis, for example, may yield an entirely different picture of subjects' accuracy, bias, and consistency than would a horizontally imposed axis. The methods offered herein also involve the imposition of arbitrary axes; however, the pair of orthogonal axes suggested serve merely as a frame of reference for describing subjects' 2-D performance. The coordinates of subjects' individual shots, subjects' centroids, and group centroids are certainly dependent upon the choice of axes, but all resulting measures and analyses involving accuracy, bias, and consistency are entirely independent of that choice.

The invariance of the proposed methods across different choices of axes is predicated upon a key, and we hope obvious, assumption. This is that from one 2-D trial to the next, from one subject to the next, and from one group to the next, the researcher uses the same axes with the same sized units. Of course, this proviso already accompanies work in 1-D performance tasks. Comparing
performance measures on different scales is virtually without meaning. So, in sum, the steps the researcher should follow are:

1. Impose a set of perpendicular axes on your 2-D performance space, letting the origin (0,0) correspond to the center of the target. The units on both axes should be identical, and meaningful to the researcher. Use these axes consistently across all trials, subjects, and groups.

2. For each shot made by each subject, measure the coordinates as precisely as possible, using the imposed axes.

3. Carefully examine the research hypotheses and match them to the appropriate measures and methods of analysis presented in this paper.

We hope that, by following the procedures outlined in this article, the researcher will be able to conduct more meaningful analyses for 2-D performance tasks.

REFERENCES


Submitted June 16, 1993
Revised December 14, 1993