The Relation of Self-Efficacy Measures to Sport Performance: A Meta-Analytic Review

Sandra E. Moritz, Deborah L. Feltz, Kyle R. Fahrbach, and Diane E. Mack

This meta-analysis examined the relationship between self-efficacy and performance in sport. Based on 45 studies (102 correlations), the average correlation between self-efficacy and sport performance was .38. Given the heterogeneity of findings, follow-up univariate and multivariate moderator analyses were conducted. Results indicated that the most important moderator was concordance, thereby highlighting the importance of matching the self-efficacy and performance measures. Additional moderators we examined included the types of self-efficacy measures, the types of performance measures, the nature of the task, and the time of assessments. These variables accounted for approximately 44% of the variance in the self-efficacy-performance relationship. Practical implications of the findings are discussed.

Key words: measurement

The construct of self-efficacy has provided the impetus for research studies across a number of domains. Self-efficacy describes the belief one has in being able to execute a specific task to obtain a certain outcome (Bandura, 1997). It is not concerned with the skills an individual has but rather with the judgments of what one can do with whatever skills he or she possesses. Self-efficacy, then, can be considered a situationally specific self-confidence (Feltz, 1988a). Self-efficacy is theorized to influence the activities individuals choose to approach, the effort they expend on such activities, and the degree of persistence they demonstrate in the face of failure or aversive stimuli (cf. Bandura, 1997). More specifically, the greater the efficacy, the greater the pursuit of challenge, and the higher the goal striving.

Many researchers have examined the relationship between self-efficacy and performance in sport. In our literature review, we found that correlations between self-efficacy and performance ranged from a high of .79 (e.g., Martin & Gill, 1991) to a low of .01 (e.g., McAuley, 1985a), and in some cases the correlations were negative (e.g., McCullagh, 1987). The magnitude and direction of the relationship between self-efficacy and performance varies considerably. Bandura (1986, 1997) stated that although self-efficacy judgments are functionally related to action, a number of factors can affect the strength of the relationship. In this review, we have summarized the literature pertaining to the relationship between self-efficacy and performance in sport via meta-analytic techniques. The reasons for conducting this meta-analysis were to clarify the existing literature and provide recommendations for researchers interested in assessing self-efficacy. In the following sections we review how self-efficacy has been assessed and how performance has been measured. We next describe the moderators that may affect the relationship.

Submitted: October 23, 1997
Accepted: March 16, 2000

Sandra E. Moritz is with the Department of Physical Education and Exercise Science at the University of North Dakota. Deborah L. Feltz is with the Department of Kinesiology at Michigan State University. Kyle R. Fahrbach is with the Department of Measurement and Quantitative Methods at Michigan State University. Diane E. Mack is with the Department of Physical Education, Recreation and Leisure Studies at Brock University.
between self-efficacy and performance. These moderators include (a) type of assessment of self-efficacy and performance, (b) concordance, (c) the participant's experience with task, and (d) time of assessments.

**Self-Efficacy Measures**

Bandura (1997) noted that disparities in observed relationships between self-efficacy beliefs and action might stem from assessment deficiencies. He advocated a micro-analytic approach that involved measuring efficacy in terms of particularized judgments of capability that vary across realms of activity, under different levels of task demands within a given activity domain, and under different situational circumstances. Accordingly, self-efficacy measures are usually constructed by listing a series of tasks, often varying in difficulty, complexity, stressfulness, or some other dimension depending on the particular function being explored (Bandura, 1982). In this approach, individuals are asked whether they can perform at specific levels for a specific task (responses are either "yes" or "no"). Then they provide the degree of confidence (or strength of their response) for those items designated as "yes" (usually rated on a 100-point probability scale from total uncertainty to total certainty) for each level.

Other methods have been used to assess self-efficacy, although these measures do not correspond to Bandura's recommendations for assessment. For example, several studies have used a single-item format to assess self-efficacy (e.g., Weinberg, Yukelson, & Jackson, 1980; Wells, Collins, & Hale, 1993). One problem with single-item measures, however, is the possibility that respondents will interpret the items as an assessment of their confidence in outcome expectancies (Lee & Bobko, 1994). Outcome expectancies do not assess judgments of capability to accomplish a certain performance level but rather the consequences that a behavior may produce (Bandura, 1986). Outcome expectancies are usually not as predictive of performance as are efficacy expectations (Bandura, 1997).

Other constructs have been used to describe efficaciousness in sport, and measures have been developed to accompany these constructs. Although we consider these measures to be more general when compared with the task-specific self-efficacy measures, we have included them in our meta-analysis because they fit our criteria as a self-efficacy measure. That is, they assess the cognitive process by which people make judgments about their capabilities to accomplish a particular goal in a sport or motor performance context (Feltz & Chase, 1998). As Feltz and Chase stated, "That goal might be quite narrow (e.g., bunting a ball down the third baseline) or more broadly defined (e.g., performing successfully in one's sport)" (p. 65). Also, common among all the measures in our meta-analysis is the definition of self-efficacy as a judgment about what one thinks one can do not the skills one has. These general or domain-specific measures include the self-confidence subscale of the Competitive State Anxiety Inventory (CSAI-2; Martens, Burton, Vealey, Bump, & Smith, 1990) and the State Sport Confidence Inventory (SSCI; Vealey, 1986).

**Performance Measures**

How performance is assessed in research studies may also be important to the self-efficacy-performance relationship. According to Bandura (1997), the performance part of the relationship must also be adequately assessed under appropriate circumstances (e.g., as little measurement error as possible, no ambiguity of task demands). Errors of performance assessment place an upper limit on how highly efficacy beliefs can correlate with performance.

Bandura (1997) defined performance as an accomplishment that must be specified by descriptive markers, such as the number of serves in bounds in tennis; first, second, third, etc., place in a competition; or a judge's numerical rating of the accomplishment. In our review, we found that performance measures could be classified as subjective, objective, or self-report. Bandura referred to subjective performance measures as "social judgments" characterized by situations when a participant's performance is rated by an external observer or coach (e.g., figure skating, gymnastics, diving, snowboarding). Although evaluators may follow guidelines for performance assessment, ultimately the performance score in these circumstances is subjectively determined. Objective performance measures, on the other hand, are generally quantitative and typically more outcome-based than the subjective performance measures (e.g., wins, points scored). Self-report performance measures, although few in our review, include such measures as effort ratings and recall items. Effort ratings can also be accomplishments, if the objective of the task is to exert effort.

**Moderators of the Self-Efficacy and Performance Relationship**

**Type of Assessment of Self-Efficacy and Performance**. There is sufficient research and theory to suggest that measuring self-efficacy as a global trait (i.e., general self-efficacy measures) or using proxy measures of performance (e.g., people's reports of what they do or the ratings of others) may affect the relationship between self-efficacy and performance (Bandura, 1997; Harrison, Ranier, Hochwarter, & Thompson, 1997; Stajkovic & Luthans, 1998). Therefore, we hypothesized that (a) correlations would be lower for those studies using general and single-item measures of self-efficacy (compared to task-specific self-efficacy measures), and (b) correlations would be lower for those studies using self-report or subjective performance measures (compared to the objective performance measures).

**Concordance Between Measures**. Evidence suggests that in addition to the type of self-efficacy and performance measure another variable affecting the relationship between
these variables may be the concordance between the measures. There are a number of cases in which self-efficacy and performance are appropriately assessed, but they relate to different competencies (cf. Stajkovic & Luthans, 1998). Bandura (1997) stated, "The structure of the relationship between efficacy beliefs and action requires that both tap similar capabilities" (p. 62). This quote implies that the self-efficacy assessment and the performance criteria must match, even when one is aggregating self-efficacy items to conduct a correlational analysis with performance. This match is what we have referred to as concordance between assessments, and it is our third hypothesis. That is, if the self-efficacy measure and performance measures are not concordant, then one should expect a lower correlation between these constructs because the self-efficacy and performance measures would assess different types of capabilities. A lack of concordance would be evident if a researcher assessed tennis skills micro-analytically (i.e., assessed an individual's confidence in his or her backhand, forehand, and serve) but then relied on the individual's overall win-loss record as the performance measure. The point is that studies that mismatch measurement levels may result in lower levels of relationships.

Nature of the Task. The third factor we believe will moderate the self-efficacy-performance relationship is the nature of the task used in the research study. In sport studies, investigators use tasks new to the participants (i.e., novel tasks) or tasks the participants are familiar with. Self-efficacy beliefs are a major determinant of behavior only when people possess the requisite subskills (Bandura, 1986). Discrepancies between efficacy beliefs and performance are most likely to occur when one has little information on which to base efficacy judgments, such as when one is learning a skill (Bandura, 1977). This does not mean that self-efficacy has no influence in the absence of skills. On the contrary, people's beliefs in their ability to learn can influence skill development. However, we expect the correlations between self-efficacy and performance to be lower when participants are performing novel tasks as compared to experienced participants.

Time of Assessments. Postperformance self-efficacy may be important to acknowledge, but it has received limited empirical attention. Bandura (1986) proposed a reciprocal relationship between self-efficacy and performance; whereby self-efficacy influences performance, and performance influences subsequent self-efficacy. Typically in assessing this reciprocal relationship, self-efficacy judgments after one trial are made for the next trial. Feltz (1982) and McAuley (1985b) found that the correlation between sport-skill performance and subsequent self-efficacy scores was stronger than the correlation between self-efficacy scores and subsequent performance scores. In their meta-analysis, Multon, Brown and Lent (1991) found that correlations between self-efficacy and academic achievement estimated from posttreatment measures of self-efficacy were significantly greater than those obtained from pretreatment measures. This finding supports Bandura's (1997) contention that performance is a powerful source of efficacy but that efficacy judgments are just one determinant of sports behavior. Similar to the other meta-analysts, we included this variable and hypothesize the same result.

To summarize, we believe that a number of factors may affect the relationship between self-efficacy and performance in sport. Our review of self-efficacy theory and research allowed us to propose a number of directional hypotheses. The first two hypotheses—higher correlations would be evident in those studies that used (a) task-specific self-efficacy measures, and (b) objective performance measures—are related to issues of assessment (cf. Stajkovic & Luthans, 1998). Our third hypothesis was that higher correlations would be obtained when the measures were concordant. We also proposed a fourth hypothesis—that the relationship between self-efficacy and performance would be greater when the participants were required to complete a task with which they were more familiar (compared with more novel tasks). Our last hypothesis relates to the order between the measures. We hypothesized that postperformance self-efficacy measures would correlate more strongly than preperformance self-efficacy measures.

Method

Study Selection

We used three techniques to locate studies for the meta-analysis. First, computer searches of SportDiscus, PsychLit, Sociolite, and Educational Resources Information Center (ERIC) were conducted using self-efficacy, self-confidence, and performance as the qualifying terms. Second, the reference lists of all articles obtained were examined for other relevant studies. Third, the tables of contents of 12 journals were searched for additional studies (i.e., Canadian Journal of Applied Sport Science, Exercise and Sport Sciences Reviews; International Journal of Sports Medicine; International Journal of Sport Psychology; Journal of Applied Sport Psychology; Journal of Sport Behavior; Journal of Sport and Exercise Psychology; Journal of Teaching in Physical Education; Perceptual and Motor Skills; Research Quarterly For Exercise and Sport; The Sport Psychologist; Therapeutic Recreation Journal).

These procedures produced an initial sample of 263 published papers. To be included in the meta-analysis, each study had to: (a) provide a measure of self-efficacy; (b) provide a measure of performance; (c) provide a correlation between self-efficacy and performance; and (d) be related to sport rather than exercise or physical activity. In addition, we required that the studies involved participants who had a mean age of over 15 years. This requirement was used, because developmental differences
in children’s capacity to differentiate between ability and effort may influence the reliability and validity of self-efficacy measures for younger populations (see Feltz & Chase, 1998). More specifically, there is less perceptual accuracy of abilities at younger ages (see Chase, Ewing, Lirgg & George, 1994). It was beyond the scope of this meta-analysis to examine all the developmental issues surrounding self-efficacy measures in children.

Application of these criteria resulted in rejecting 218 studies. The most common reasons for rejecting studies were the lack of self-efficacy measure \( (n = 24) \), the lack of a performance measure \( (n = 39) \), or the failure to report a correlation between self-efficacy and performance \( (n = 60) \). Eighteen studies were rejected because they did not meet the minimum age requirement for inclusion. An additional 22 articles were rejected because they were not research studies, and 40 were rejected because they were not related to sport. Other reasons for rejecting studies included being a single-subject design \( (n = 4) \), reporting only a point biserial correlation \( (n = 5) \), having participants evaluate someone else’s self-efficacy \( (n = 1) \), investigating coaching efficacy \( (n = 2) \), or being a meta-analysis \( (n = 1) \). We rejected two studies because the same data were reported elsewhere and were already included in the meta-analysis. Overall, the 45 studies that met inclusion criteria yielded 102 correlations.

All the studies in this review reported a Pearson product-moment correlation coefficient between self-efficacy and performance. We used electronic mail to obtain missing information from a number of authors. The studies included in our meta-analysis included 3,055 participants. The studies ranged in sample size from 7 to 216 \( (M = 67.89, \text{ Median} = 60, \text{ Mode} = 40, \text{ SD} = 41.25) \). The samples were composed of athletes \( (n = 1,295, 42\%) \), undergraduate nonathletes \( (n = 959, 31\%) \), and others \( (n = 801, 26\%) \).

Variables Coded From Each Study

We coded: (a) correlation between self-efficacy and performance, (b) type of self-efficacy measure (task-specific, domain-specific, single-item), (c) type of performance measure (subjective, objective, self-report), (d) concordance between self-efficacy and performance measures (concordant, not concordant), (e) nature of the task (novel, familiar), and (f) time of the self-efficacy assessment in relation to performance (before, after).

Dependencies Between Study Correlations

In analyzing studies for the meta-analysis, it was apparent that some studies provided more than one correlation between self-efficacy and performance. When more than one self-efficacy measure (e.g., a task-specific measure and the SSCI) was correlated with the same performance measure, both the correlations were coded. On the other hand, if the same self-efficacy measure was correlated with multiple performance measures, depending on the type of performance measures, we averaged the correlations. For example, if a single study reported separate correlations between self-efficacy and two performance measures and if both of the performance measures were objective, then we averaged the correlations. However, if one of the performance measures was subjective and the other was objective, then both correlations were retained.

When averaging correlations, we paid particular attention to the issue of concordance between measures. That is, if a study used a task-specific measure of self-efficacy and assessed performance in the same manner (i.e., assessed self-efficacy for the number of golf putts made and measured the number of putts made), then this correlation was concordant. If a study used a more general measure of self-efficacy (such as the CSAI-2 or the SSCI) and assessed performance in a specific manner (i.e., number of golf putts made) then the correlation was recorded as not concordant.

Another issue in combining correlations resulted when multiple self-efficacy and performance assessments were reported in a single study. For example, if a study reported a correlation between self-efficacy and performance for a number of time periods or trials, we averaged the correlations across trials (cf. Gully, Devine, & Whitney, 1995; Wolfe, 1986). However, if participants were exposed to an intervention (e.g., goal-setting, imagery) between the self-efficacy and performance assessments, we used only the pre-intervention correlations. Finally, if a study used path analysis techniques, we used only the first “wave” in our meta-analysis.

A major concern for meta-analysts pertains to the methods of combining data (Strube, 1985). The methods we used did not preclude an investigation into the differences and similarities of results for different categories. All studies were coded independently by two of the authors, and their ratings were compared. In addition, a statistical consultant checked all the variable combinations and codings. Agreement rates were between 95 and 100%, depending on the variable. Disagreements were resolved by discussion.

Although the aforementioned methods of combining multiple correlations from the same study have been used in previous meta-analyses (e.g., Carron, Hausenblas, & Mack, 1996; Gully et al., 1995; Hausenblas, Carron, & Mack, 1997), they do not completely eliminate the possibility of generating multiple correlations from a single study. Debate regarding the number of correlations that should be used from one study in a meta-analysis is considerable (cf. Wolfe, 1986). Using multiple correlations from a single study or calculating effect size estimates from identical groups of participants violates the assumption of independence of correlations (Hedges, 1994). This dependence may inflate the Type I error rate for a meta-analysis and may result in erroneous characterizations of the between-studies variation in effects. Previous meta-analysts (e.g.,
Moritz, Feltz, Fahrbach, and Mack

Carron et al., 1995; Hausenblas et al., 1997; Kyllo & Landers, 1995) have ignored this issue, and it is likely that their resulting estimates are inaccurate (Hedges, 1994).

To mitigate the effect of correlation dependency, we used a generalized least squares (GLS) approach to our meta-analysis (Becker, 1992). The GLS method assumes the correlations from different samples are independent, but that correlations from the same sample can be dependent. The dependencies among the within-sample correlations are accounted for in the variance-covariance matrix whose inverse leads to the weight that a particular correlation receives. When two correlations arise from independent samples, their covariance is zero. However, when two correlations are derived from the same sample, they covary as the correlations themselves are actually correlated. This covariance can be modeled (cf. Olkin & Saotani, 1976). For example, suppose a single study used two measures of performance. If one knows the correlation between the performance measures, one could estimate the covariance between the correlation of self-efficacy and the first performance measure and the correlation between self-efficacy and the second performance measure. Similarly, if a study reported two measures of self-efficacy, the correlation between the self-efficacy measures could be used to calculate the necessary covariance. Thus, for those studies that had multiple self-efficacy or performance measures, we coded not only the correlation between self-efficacy and performance but also the correlations among the performance and self-efficacy measures.

The advantages of the GLS procedure are notable. Similar to analysis of variance (ANOVA) procedures, the assumption of independence between observations is not robust in the typically used weighted least squares (WLS) meta-analysis procedure. The repercussions are that the weightings of the correlations would be incorrect. In addition, although the estimates of the mean are unbiased, the confidence intervals are in error (i.e., generally narrower than they truly should be).

**Computation and Analyses of Correlations**

The effect size metric was $r$, the correlation between self-efficacy and performance. To normalize these values, all the correlations were changed to $z$ scores using the Fisher "$r$ to $z$" transformation. A weighted estimate of the common correlation was computed using formulas similar to those found in Becker (1992). The key difference between the procedure used in this analysis and what is usually done in a WLS analysis is that the weighting matrix (the inverse of the variance-covariance matrix) had some off-diagonal values that were not zero. The mean correlation and the upper 97.5 and lower 2.5 percentiles of Fisher’s $z$ were then converted back into the $r$ matrix. The estimation of the population correlations and the 95% confidence intervals for the self-efficacy estimates are, therefore, completely in terms of $r$.

To determine if the correlations were representative of a single population correlation, we employed Hedges and Olkin's (1985) test of homogeneity ($H$). A nonsignificant $H$ would indicate that the correlations were homogenous, and no additional moderator analysis would be required. A significant $H$ indicates heterogeneity among the correlations, and additional moderator tests are warranted. Follow-up testing investigated whether potential moderating variables could individually (i.e., univariate analyses) or in groups (i.e., multivariate moderator analysis) explain any of the surplus variation (heterogeneity) that existed in the sample correlations.

To elaborate, just as one can run both univariate and multivariate analyses of variables in more standard settings in the social sciences, one can do the same in meta-analysis. In the more standard settings, tests and correlations are used to examine univariate relationships, while multiple regression analyses are used to examine the partial effect the variables have on the outcome after there have been controls for all other variables. The equivalent of a multiple regression analysis in a meta-analysis is a multiple moderator analysis (MMA).

The univariate procedure is analogous to an ANOVA in which between-groups and within-groups variation is separated. Here, the between-group sum of squares ($H_n$) was tested to determine whether the moderators tested were significantly related to the size of the sample correlation. The within-group sum of squares ($H_s$) was used to test the hypothesis that, after accounting for the moderator(s) in question, the sample correlations seemed to derive from a homogenous single population correlation.

All the studies were partitioned into a number of different categories based on the potential moderators. The ANOVA-like comparisons conformed to the following relationship: $H_n = H_k - H_s$. Both of the values for $H_n$ and $H_k$ are distributed as a $\chi^2$, differing only in their respective degrees of freedom. $H_n$ has $p-1$ degrees of freedom, where $p$ is the number of categories; and $H_k$ has $k(p-1)$ degrees of freedom, where $k$ is the total number of samples, and $p$ is as previously defined. A significant $H_n$ suggests the correlations differed across the moderator categories and that the moderator affected the self-efficacy-performance relationship. A significant $H_k$ coupled with a nonsignificant $H_s$ indicated that the moderator provided an adequate model of correlation variability, because the correlations differed across categories and were homogenous within categories. A significant $H_k$ and a significant $H_s$ indicated that the moderator explained some variance, but the correlations still remained heterogeneous within categories.

With respect to the MMA, the interpretation of the coefficients is very similar to what it would be in a multiple regression; each slope represents the effect of a one-unit change in the moderator on the outcome, when all other predictors are held constant. The $p$ values and substantive conclusions may differ depending on whether univariate or multivariate analyses are used, as different
hypotheses are being tested for each variable. For instance, in the univariate analysis, we test the hypothesis that the type of self-efficacy measure is related to the correlation under investigation between self-efficacy and performance. In the MMA we test the hypothesis that the type of self-efficacy measure, controlling for all other potential moderators (such as whether the measure was concordant and when the assessment was conducted), is related to the correlation between self-efficacy and performance. It is important to note that if there is a relationship between the type of self-efficacy measure and other moderators, then the tests of the univariate and multivariate hypotheses can lead to different conclusions.

File Drawer Analysis

A number of meta-analysts have discussed the issue of publication bias. In many cases, the decision to publish is influenced by the presence or absence of a statistically significant effect, whereby significant results are more likely to be published. Beggs (1994) stated that if a meta-analysis is restricted to published studies there is a risk it will lead to biased conclusions. Based on the large range of correlations we obtained between self-efficacy and performance in this meta-analysis (.17 to .79) and the large number of correlations we used in the analysis, we felt that publication bias would not be a serious issue. However, we still calculated a “file-drawer” statistic. The file-drawer statistic refers to how many unlocated studies with null effects would have to be found to reduce the correlation to zero or to nonsignificance (cf. Orwin, 1983; Rosenthal, 1979; Wolfe, 1986). We calculated an overall fail-safe N to determine the number of studies needed to reduce the p value of the test (that the mean population correlation between self-efficacy and performance is zero) to a critical value of .001. The fail-safe N was computed to be 657. Beggs (1994) suggested this large value makes it quite unlikely that the population correlation coefficient is zero.

Results

The average correlation obtained between self-efficacy and performance in sport was .38 (95% confidence interval: .35–.41), which was significant (z = 25.80, p < .001). This correlation of .38 is considered moderate in size and corroborates what other nonsport researchers have found between self-efficacy and performance. For example, Muto et al. (1991) reported an aggregate correlation of .38 between self-efficacy and academic performance and .34 for self-efficacy and academic persistence, and Stajkovic and Luthans (1998) reported an aggregate correlation of .38 in their meta-analysis of self-efficacy and work-related performance. Although we found an aggregate correlation of .38 for self-efficacy and sport performance, the homogeneity statistic (H₁ = 446.00) revealed significant heterogeneity among correlations. In convention with other meta-analysts, we report the univariate analyses first, followed by the multiple moderator analysis (cf. Stajkovic & Luthans, 1998).

Univariate Analyses

Table 1 contains an overview of the results of the univariate analyses (i.e., the number of studies in each analysis, the aggregate correlation for the studies with 95% confidence intervals, the associated homogeneity statistics [H₀ and H₁], and the proportion of variance possible to explain). The formula used to obtain this value was H₀/ H₁-k). Note that the H₂ value differs for one individual moderator analysis (i.e., nature of the task). This difference in values reflects the differences in the number of studies included in this analysis—not all the correlations were used due to missing data.

Self-Efficacy Measures. When the studies were grouped according to type of self-efficacy measure, the correlations for each group differed significantly. Specifically, as predicted, the task-specific measures had the largest correlations with performance (r = .38). The mean correlations for the categories of domain-specific measures and single-item measures were .26 and .28, respectively. The task-specific measures were found to correlate more strongly with performance compared with the other assessment methods, including the SSCI (r = .28), which is the instrument for Vealey’s (1986) conceptualization of sport confidence, and the self-confidence subscale of Martens et al. (1990) CSAI-2 (r = .24).

Performance Measures. The results of the analysis using types of performance measures (i.e., subjective, objective, self-report) as a moderator indicated that the largest correlations were obtained for those studies which subjectively assessed performance (r = .47), followed by self-report (r = .44) and objective measures (r = .34).

Concordance Between Measures. Consistent with our predictions, we found higher correlations (r = .43) between self-efficacy and performance for those studies that were concordant in assessment methods compared with nonconcordant studies (r = .26). This particular moderator analysis also accounted for the greatest amount of variance—approximately 13%.

Nature of the Task. We also divided the studies into groups based on the participants’ experience with the task. Larger correlations were found for those studies that used performance tasks familiar to the participants (r = .39), compared with novel tasks (r = .31).

Time of Assessments. We divided time of assessment into before- and after-performance categories. As hypothesized, we found larger correlations for those studies that assessed self-efficacy after performance (r = .39) compared with the studies that assessed self-efficacy before performance (r = .36), although these differences were slight.
Multiple Moderator Analysis

The results for the MMA are presented in Table 2. The purpose of the MMA was to assess how much variation in the self-efficacy-performance relationship remained after accounting for the moderators and to determine which moderators significantly contributed to the self-efficacy-performance relationship when all moderators were considered simultaneously (Hedges, 1994; Hedges & Olkin, 1985).

The procedure for the MMA is analogous to a multiple regression analysis, and interpreting the multiple moderator analysis is akin to interpreting multiple regression equations that have categorical data. The dependent variable was the correlation between self-efficacy and performance. Similar to other regression analyses, the MMA calculates an overall intercept value and individual beta weights. The individual beta weights reflect how much higher or lower the correlation between self-efficacy and performance would be for each level of the dummy-coded variables using the intercept value as the reference category. Therefore, the individual beta weights can be used to determine the effect of an individual moderator when all other moderators are held constant; the intercept in the MMA assumes the beta weights for the first “level” of all the dummy-coded moderators. Thus, the regression equation is interpretable only when one considers all the moderators that are included in the equation. Because of missing data, only 100 of the 102 correlations were used.

To determine the amount of variance in the criterion all six variables were able to account for, we divided $H_b$.

Table 1. Summary of univariate analyses

<table>
<thead>
<tr>
<th>Category</th>
<th>$K$</th>
<th>$r$</th>
<th>95% Confidence interval</th>
<th>% Variance accounted for</th>
<th>Homogeneity test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample</td>
<td>102</td>
<td>.38</td>
<td>.35/41</td>
<td></td>
<td>$H_t = 446.00^{**}$</td>
</tr>
<tr>
<td><strong>Self-efficacy measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total variation</td>
<td>102</td>
<td></td>
<td>2.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-group self-efficacy measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-group self-efficacy measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task-Specific</td>
<td>84</td>
<td>.38</td>
<td>.35/40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain-specific</td>
<td>13</td>
<td>.26</td>
<td>.18/34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single items</td>
<td>5</td>
<td>.28</td>
<td>.17/39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total variation</td>
<td>102</td>
<td></td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-group performance measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-group performance measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td>84</td>
<td>.34</td>
<td>.32/37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective</td>
<td>16</td>
<td>.47</td>
<td>.41/53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-report</td>
<td>2</td>
<td>.44</td>
<td>.30/56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concordance between measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total variation</td>
<td>102</td>
<td></td>
<td>12.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-group concordance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-group concordance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concordant</td>
<td>53</td>
<td>.43</td>
<td>.40/46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not concordant</td>
<td>49</td>
<td>.26</td>
<td>.22/30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature of the task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total variation</td>
<td>101</td>
<td></td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-group task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-group task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel</td>
<td>35</td>
<td>.31</td>
<td>.26/.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar</td>
<td>66</td>
<td>.39</td>
<td>.36/.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of assessments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total variation</td>
<td>102</td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-group time of assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-group time of assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before</td>
<td>84</td>
<td>.36</td>
<td>.33/.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>After</td>
<td>18</td>
<td>.39</td>
<td>.35/.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.

**p < .01.
Thus, together these variables accounted for 22% of the variance in self-efficacy-performance correlations. This value may be slightly conservative, considering that some of the remaining variance is attributable to sampling error. More specifically, to determine the amount of variance that could be attributed to sampling error we calculated (k−1), where k is the number of correlations used in this analysis (k = 100), and divided this value by H². Thus, an additional 22% of the variance may be attributable to sampling error. When these values were added together, we were able to account for approximately 44% of the variance in the self-efficacy performance relationships. This value suggests there are other variables that may continue to moderate this relationship.

In both the univariate and multiple moderator analyses, concordance was the most important variable. This variable accounted for approximately 13% of the variance between self-efficacy and performance in the univariate analyses and remained statistically significant in the MMA. The second and third most important variables in the univariate analyses were the type of self-efficacy measure and the type of performance measure. Interestingly, neither of these variables was statistically significant in the MMA. Conversely, the nature of the task and time of assessments accounted for the least variance according to the univariate analyses, but they were statistically significant in the MMA. We comment further on these results in the discussion.

## Discussion

The advantage of conducting a meta-analysis compared to other more qualitative reviews is that it provides an estimate of the mean correlation of the strength of a relationship, given a whole population of studies. In this case, the correlation between self-efficacy and performance was .38. This value suggests that self-efficacy beliefs have a positive and moderate relationship with performance in sport. The studies we reviewed differed widely in the self-efficacy and performance measures used as well as in other characteristics. Because of this diversity, we believe that the results of our meta-analysis have considerable generalizability within the sport domain.

A purpose of this meta-analysis was to clarify the existing literature pertaining to the relationship between self-efficacy and performance in the sport setting. Because the correlations between self-efficacy and performance were not homogenous, we were able to separate a number of variables of interest and to investigate them as potential moderators. We determined that a number of study features had a modifying effect on the self-efficacy-performance relationship, and although our analysis of potential moderating variables did not yield completely adequate models (i.e., heterogeneity remained within groupings), the analysis did produce several findings relevant to self-efficacy assessments.

Based on self-efficacy theory and previous research, we posed a number of directional hypotheses. The first hypothesis dealt with the type of assessment of self-efficacy and performance. We expected task-specific self-efficacy measures to correlate most strongly with performance compared to the other assessment methods. The results from the univariate analysis supported our hypothesis. That is, those scales that measured self-efficacy for the performance of a specific task resulted in the largest correlations with performance compared with all other measures (i.e., general measures, single items).

Also related to the type of assessment of self-efficacy and performance was the consideration of the performance measure. Contrary to our hypothesis, we found that those studies with a subjective performance measure (e.g., external ratings) had the highest aggregate correlation with performance when compared to objective ratings or self-report ratings in the univariate analysis. However, the relationship between self-efficacy and performance held up well in all circumstances.

The finding that the self-efficacy-performance relationship was so strong with subjectively determined performance measures was surprising to us. One may expect that subjective performance measures have lower reliabil-

## Table 2. Multiple moderator analysis

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Beta</th>
<th>95% Confidence interval</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.50</td>
<td>.44/.57</td>
<td>.00**</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task-specific</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain-specific</td>
<td>-.05</td>
<td>-.15/.05</td>
<td>.33</td>
</tr>
<tr>
<td>Single items</td>
<td>-.05</td>
<td>-.18/.08</td>
<td>.45</td>
</tr>
<tr>
<td>Performance measure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective</td>
<td>.08</td>
<td>-.01/.18</td>
<td>.09</td>
</tr>
<tr>
<td>Self-report</td>
<td>-.04</td>
<td>-.20/.13</td>
<td>.67</td>
</tr>
<tr>
<td>Concordance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concordant</td>
<td>-.26</td>
<td>-.33/.19</td>
<td>.00**</td>
</tr>
<tr>
<td>Not concordant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature of the task</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar</td>
<td>-.19</td>
<td>-.26/.12</td>
<td>.00**</td>
</tr>
<tr>
<td>Novel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before performance</td>
<td>.05</td>
<td>.01/.09</td>
<td>.03*</td>
</tr>
<tr>
<td>After performance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Betas and confidence values are expressed in terms of z; the formula to convert z to r is: r = e^z - 1/ e^z + 1.

*p < .05

**p < .01
ity than other performance measures thereby attenuating the $r$s for subjective measures. This lower reliability may be due to a lack of consensus among the judges as to what constitutes a good performance. The studies in our meta-analysis that used subjective performance measures (i.e., Barling & Abel, 1983; Feltz, 1982, 1988b; Feltz, Landers, & Raeder, 1979; Feltz & Mugno, 1983; Maynard & Howe, 1987; McAuley, 1985b; McAuley & Gill, 1983; Treasure, Monson, & Lox, 1996; Vealey, 1986) involved evaluators ranging between 1 and 3 ($M = 1.63, SD = .74$). For the studies that reported interrater reliability coefficients, the values ranged from .91 to 1.0. Perhaps the low number of evaluators and the high interrater reliability values negated the variability one may expect among judges’ expectations. This may have contributed to the high correlations between self-efficacy and performance. In addition, Bandura (1997), in commenting on subjectively judged performances, stated, “The accuracy of efficacy beliefs will depend partly on knowledge of the subjective criteria on which one’s performance will be judged” (p. 65). It is conceivable that in these studies there was a high consensus among athletes and judges regarding what constituted a good performance.

Another explanation for this finding may be related to the nature of the objective performance measures. A number of researchers (e.g., Harrison et al., 1997) have suggested that outcome variables are not good measures of performance, because they measure not only an individual’s behavior but also reflect the occurrences in the environment over which an individual has little control. It may be that subjective measures are more sensitive performance variables than “outcome” variables. On the other hand, an individual has little control over a number of things that may also influence subjective performance measures (e.g., appearance). Future researchers may want to examine the relationship between self-efficacy and objective and subjective performance measures in greater detail.

One last possible explanation relates to the concordance between the measures. The studies that used subjective performance measures were also found to be more concordant with the self-efficacy measure than those using objective performance measures. Of the 16 correlations using subjective performance measures, only 3 were not concordant. Thus, the strength of this relationship may be more determined by the concordance than by the performance measures per se.

Our third hypothesis was that higher correlations between self-efficacy and performance would be obtained when the measures are concordant. Our meta-analysis supports our hypothesis. Of all the coding characteristics examined, by far the most important was concordance. As an individual moderator, concordance accounted for the largest proportion of the variance in the correlations between self-efficacy and performance. Even when we controlled for other variables, concordance remained statistically significant in the MMA. The results from these analyses showed that when nonconcordant measures were used, the resulting correlation between self-efficacy and performance dramatically decreased.

It is worthwhile to elaborate on the importance of concordance and how the significance of this moderator affected the other moderators considered in this meta-analysis. Interestingly, despite the significant univariate findings, neither the type of self-efficacy nor performance measures was significant in the MMA. Given these conflicting results, a reasonable question pertains as to which results should take precedence for interpretation.

One reason the results from the univariate analyses differ from the MMA is that different hypotheses are being tested. In the univariate analyses we tested the hypothesis that a moderator was related to the correlation between self-efficacy and performance. In the MMA we tested the hypothesis that the moderator, controlling for all other potential moderators (such as whether the measure was concordant and when the assessment was conducted), was related to the correlation between self-efficacy and performance. Table 3 shows the relationships between some of the moderators. Thus, the tests of the univariate and multivariate analyses show discrepancies in results likely due to the presence of some multicollinearity in the data.

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Self-efficacy measures (domain-specific)</td>
<td></td>
<td></td>
<td>-.09</td>
<td>.02</td>
<td>.09</td>
<td>-.06</td>
<td>.39</td>
</tr>
<tr>
<td>(2) Self-efficacy measures (single item)</td>
<td>-.09</td>
<td></td>
<td></td>
<td>-.14</td>
<td>-.10</td>
<td>-.03</td>
<td>-.13</td>
</tr>
<tr>
<td>(3) Time of assessment (after)</td>
<td>-.02</td>
<td>.14</td>
<td></td>
<td></td>
<td>.18</td>
<td>-.06</td>
<td>-.02</td>
</tr>
<tr>
<td>(4) Performance measures (subjective)</td>
<td>.09</td>
<td>-.10</td>
<td></td>
<td>.18</td>
<td></td>
<td>-.06</td>
<td>-.24</td>
</tr>
<tr>
<td>(5) Performance measures (self-report)</td>
<td>-.06</td>
<td>-.03</td>
<td>-.06</td>
<td></td>
<td>-.06</td>
<td></td>
<td>-.14</td>
</tr>
<tr>
<td>(6) Concordance (not concordant)</td>
<td>.39</td>
<td>-.13</td>
<td>-.02</td>
<td>-.24</td>
<td>-.14</td>
<td></td>
<td>-.17</td>
</tr>
<tr>
<td>(7) Nature of the task (novel)</td>
<td>-.28</td>
<td>.12</td>
<td>.11</td>
<td>.28</td>
<td>-.10</td>
<td>-.17</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The intercept in the multiple moderator analysis assumes the beta weights for the first “level” of all of the dummy-coded moderators; thus, they are not included here.
For example, in the univariate analysis, the average correlation for domain-specific self-efficacy measures was .28, which is moderate. One would expect this variable to remain statistically significant in the MMA. However, the reason the domain-specific self-efficacy measures becomes insignificant in the MMA is likely due to the very high correlation it has with concordance (r = .39). Concordance is by far the more important variable measured in terms of its relationship with the correlation between self-efficacy and performance. Once concordance is held constant, the effect of the domain-specific self-efficacy measures becomes practically nonexistent. Similarly, the subjective performance measures correlates moderately with concordance (.24). Therefore, it would be expected that the results of the hypothesis test would be different in the MMA, because concordance is controlled for in the MMA. Note, on the other hand, that the time of assessment variable is correlated very weakly with all of the variables. Its correlation with concordance is only .02. Not surprisingly, its significance and the size of the effect are about the same in the univariate analysis and the MMA. Similarly, the nature of the task is only correlated (.17) with concordance, and it is also statistically significant in both analyses.

In addition, the sample size plays a role. There are only 13 observations for domain-specific self-efficacy measures, and the effect is more easily changed after controlling for the other variables. Single item self-efficacy measures have only five cases, and the effect can change very easily after controlling for many of the other variables. The same holds true for categories corresponding to the performance measures—the subjective performance measures has only 16 cases, and the self-report performance measures has only two cases—holding more variables constant can easily change the estimate of the beta for these parameters. Nature of the task, on the other hand, has plenty of observations at both levels (n = 35 and 66), and it is less prone to be affected when holding other variables constant.

The types of self-efficacy and performance measures are both important, but more essential is the concordance between the measures. Thus, if one wishes to maximize the relationship between self-efficacy and performance, we strongly recommend using concordant measures. For example, if one is interested in using win-loss as a performance measure for the sport of tennis, then the self-efficacy measure should ask respondents how confident they are that they will win the game-set-match. Alternatively, if one is interested in an individual’s confidence in specific tennis skills (i.e., forehand, backhand, serve, lob), then the performance measure should assess these specific skills. One final thought about concordance—nonconcordance isn’t necessarily “bad,” it is just nonconcordant. That is, nonconcordance might be useful if one were concerned about generalizability.

Our fourth hypothesis addressed the nature of the task. Consistent with self-efficacy theory, correlations between self-efficacy and performance were lower when researchers employed tasks novel to their participants. The results from the MMA also support this finding: holding other variables constant, the correlation between self-efficacy and performance decreased by .19 units when researcher used novel tasks. This finding supports Bandura’s (1997) contention that accurate self-efficacy judgments require knowledge of the demands of a task. As Bandura states, “If one does not know what demands must be fulfilled in a given endeavor, one cannot accurately judge whether one has the requisite abilities to perform the task” (p. 64). With novel tasks, the task demands are more ambiguous and, thus, create greater discrepancies between efficacy judgments and performance than with well-rehearsed tasks.

Finally, we expected that correlations between self-efficacy and performance estimated from postperformance measures of self-efficacy would be significantly larger than those obtained from preperformance measures. Both the univariate analysis and the MMA results supported this hypothesis and corroborate the findings in the meta-analysis by Multon et al. (1991). Research in motor performance suggests that as a performer gains experience on a task over time performance becomes a stronger predictor of self-efficacy than self-efficacy is of performance (Feltz, 1992). These findings suggest that other variables besides self-efficacy, such as previous performances, influence the complex nature of sport performance (cf. Bandura, 1986).

Although we coded the time of self-efficacy measures relative to performance (as before and after) the specific amount of elapsed time between assessments may affect the self-efficacy—performance relationship. Bandura (1997) postulated that efficacy strength might waver as the time of performance draws near. In sport, we often strive to assess psychological constructs as close to performance as possible. It seems feasible that there is a relationship between the time assessments were taken and the resulting degree of association between self-efficacy and performance. Another possibility suggested by Bandura is that it may not be the amount of time elapsed between the measures but rather that the efficacy beliefs have been altered by intervening experiences. Researchers may wish to explore the time of assessments variable in greater detail.

This meta-analysis provides clear evidence for a significant relationship between self-efficacy and performance. The studies included used different tasks and measures. Self-efficacy is both a cause and effect of performance. Feltz (1992) suggested that where self-efficacy has not been shown to be a reliable predictor of sport performance, it probably has more to do with the way in which the constructs were measured than with the conceptual soundness of self-efficacy theory. Although self-efficacy should not be expected to fully explain human behavior (Bandura, 1986), particularly the complex behavior of sport performance (Vealey, 1986), given the demonstrated
importance of self-efficacy in enhancing performance, numerous inferences can be drawn to help individuals develop and maintain self-confidence (Feltz, 1992).

Before ending our discussion of moderator results, we must echo an important cautionary note, namely, that the conclusions we drew from our meta-analysis were somewhat compromised by the less than optimal data-reporting practices in this literature (cf. Multon et al., 1991). Many of the studies we considered for inclusion in this meta-analysis were rejected because of their failure to provide a correlation between self-efficacy and performance. Furthermore, in a number of studies, we found that seemingly basic demographic data (e.g., gender of participants) was often omitted from studies and descriptions of the procedures used were often less than adequate. Thus, we agree with other meta-analysts who implore researchers to report all results for a particular study, regardless of whether statistical significance is obtained (Multon et al., 1991; Oliver & Spokane, 1983). In addition, we request that journal editors demand these results from the authors.

A couple final points related to the heterogeneity of our findings are warranted. First, because of the large amount of heterogeneity in our findings, we believe researchers should be encouraged to conduct and publish replication studies. Second, our study was primarily focused on measurement issues. It should be evident that we did not consider a number of possible moderators. Some potential moderators include the presence or absence of feedback, the presence or absence of incentives, the presence or absence of performance constraints (i.e., the availability of or lack of necessary equipment to perform the activities), the consequences of misjudgments of efficacy, the manner in which self-efficacy judgments are made (privately or publicly), etc. Interested readers are directed to Bandura (1997) for a more detailed description of these issues.

Finally, we offer the following guidelines for researchers interested in assessing the relationship between self-efficacy and performance in sport. First, to ensure valid and reliable results, we believe that self-efficacy should be assessed in a task-specific manner and that self-efficacy strength scores should be used. Second, and perhaps most important, we recommend that self-efficacy and performance measures be concordant. Finally, we believe that postperformance as well as preperformance self-efficacy should be assessed. Along this line, researchers should test for the reciprocal effects of these variables.

References


Appendix. Studies Used in the Meta-Analysis


Notes

1. Complete coding summaries for each study used in the meta-analysis is available from the first author.
2. If an item was single-item and was task-specific, it was coded as a single item. There were only five single-item correlations (see Weinberg, Yukelson, & Jackson, 1980; Wells, Collins, & Hale, 1993; Wilkes & Summers, 1984; Woolfolk, Murphy, Gottesfeld, & Aitken, 1985). Only the Wilkes and Summers (1984) study was not task-specific (the item in their study asked for one’s confidence in “doing well”).
3. The 95% confidence intervals for each variable in the moderator analyses are asymmetrical. This asymmetry is caused by the Fisher “r to z” conversion. The confidence intervals were symmetrical when they were expressed in terms of z, but they lost symmetry when they were converted to r.

Authors’ Notes

We thank Betsy Becker for her assistance with the statistical analyses and for her comments on this manuscript. We also thank Daniel Ilgen and John Spence for their comments on an earlier version of this manuscript. Please address all correspondence concerning this article to Sandra E. Moritz, Department of Physical Education and Exercise Science, University of North Dakota, Box 8235, Grand Forks, North Dakota, U.S.A. 58202-8235.

E-mail: sandra_moritz@und.nodak.edu