

1 Running head: COLLECTIVE EFFICACY: SINGLE-ITEM MEASUREMENT

2

3

4

5 Validation of a Single-Item Stem for Collective Efficacy Measurement in Sports Teams

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

## 1 Abstract

2 The validity and reliability of an operational stem for single-item collective efficacy  
3 measurement was examined across three studies in sport. Study one used a cross-sectional  
4 design with team sports participants to assess the stem's validity by examining the ability of  
5 the Collective Efficacy Questionnaire for Sports (CEQS), Group Environment Questionnaire  
6 (GEQ), and previous performance (win %) to predict single-item collective efficacy scores.  
7 Total CEQS scores predicted single-item scores ( $\beta = .69$ ), with the CEQS Ability ( $\beta = .51$ )  
8 and Persistence ( $\beta = .15$ ) subscales the only significant predictors of single-item score.  
9 Previous performance ( $\beta = .41$ ) and three GEQ dimensions; Individual Attractions to the  
10 Group-Task ( $\beta = .18$ ), Group Integration-Task ( $\beta = .22$ ), and Group Integration-Social ( $\beta =$   
11  $.16$ ) were also significant predictors of single-item collective efficacy. Study two examined  
12 the validity and reliability of the stem using an intervention with team sports participants in a  
13 laboratory-based design. The single-item measure demonstrated high concordance (pre-  
14 intervention;  $r = .53$ , post-intervention;  $r = .73$ ) and reliability ( $r = .77, .62$ ) with CEQS  
15 scores. Study three assessed the validity and reliability of the stem using an intervention with  
16 team sports participants in a field-based design. The single-item measure reported high  
17 concordance (pre-intervention;  $r = .74$ , post-intervention;  $r = .69$ ) and reliability ( $r = .88, .87$ )  
18 with CEQS scores. The studies support the stem as valid and reliable for single-item  
19 collective efficacy measurement with team sports players.

20 *Key words:* collective efficacy, operational stem, single-item measurement, validity,  
21 reliability

22

23

24

25

## 1    **Validation of a Single-Item Stem for Collective Efficacy Measurement in Sports Teams**

2    Bandura (1982, 1997) suggested that humans often work together towards collective  
3    objectives in groups or teams and have collective efficacy beliefs regarding their functional  
4    abilities for specific tasks. Collective efficacy is defined as “a group’s shared belief in its  
5    conjoint capability to organize and execute the courses of action required to produce given  
6    levels of attainment” (Bandura, 1997, p. 477). Meta-analyses consistently indicate collective  
7    efficacy has a positive effect upon performance of groups (see e.g., Gully, Incalcaterra, Joshi,  
8    & Beaubien, 2002; Stajkovic, Lee, & Nyberg, 2009), a finding that has been replicated in  
9    both laboratory (e.g., Greenlees, Graydon, & Maynard, 2000) and field-based settings with  
10    sports teams (e.g., Myers, Feltz, & Short, 2004).

11        Collective efficacy has been conceptualized (and subsequently analyzed) both at an  
12    individual (e.g., Heuze, Sarrazin, Masiero, Raimbault, & Thomas, 2006) and group level  
13    (e.g., Gibson, 1999). Bandura (1997) advocates that each team member’s belief in the team’s  
14    overall capabilities be considered, and these individual measures aggregated to the team  
15    level. Therefore, both individual and group level approaches are suitable for use with  
16    collective efficacy measurement, with the choice of level contingent on the situation involved  
17    (i.e., suited to the specific context). Aggregated collective efficacy details a group’s overall  
18    beliefs, but does not consider individual differences within the group (Shearer, Holmes, &  
19    Mellalieu, 2009). Given that collective efficacy is ultimately measured through individual  
20    cognitions, in the current paper we adopted an individual-level approach to the manipulation,  
21    measurement, and analysis of collective efficacy perceptions.

22        A key aspect of using psychometric tools to measure social psychological constructs  
23    such as collective efficacy is the wording of the operational stem, which represents the  
24    beginning part of an item that presents the issue about which the question is asking (Roe,  
25    2008). In collective efficacy research, different stems have been used to direct the

1 participant's focus towards either their own beliefs about the team, or what they perceive the  
2 team thinks (cf. Bandura, 1997; Lindsley, Brass, & Thomas, 1995). In Short et al.'s (2002)  
3 examination of the difference between these two operational methods, the first stem assessed  
4 individual perceptions of a team's collective efficacy (*'rate your confidence that your*  
5 *team...'*), whereas the second asked the respondent to consider their team's perceptions (*'rate*  
6 *your team's confidence...'*). No significant differences were reported between the two  
7 operational stems, and consequently, the stem *'rate your team's confidence...'* was used for  
8 the Collective Efficacy Questionnaire for Sports (CEQS; Short, Sullivan, & Feltz, 2005).  
9 Short, Sullivan, et al. suggested that this stem accounts for an individual's perceptions about  
10 their team, which is different to the cognitions they experience as individuals inside/outside  
11 the team context (e.g., perceptions of their individual performance).

12 In terms of questionnaire structure, the majority of literature has measured collective  
13 efficacy using multi-item instruments that assess an athlete's confidence in their team's  
14 ability to perform significant game competencies (e.g., Myers et al., 2004). Studies have  
15 suggested that psychometric concerns regarding single-item instruments make them  
16 inappropriate for use with multidimensional constructs such as collective efficacy (see e.g.,  
17 Loo, 2002). However, empirical support does exist for the use of single-item measures,  
18 suggesting they are appropriate for use in certain circumstances (e.g., Jordan & Turner, 2008;  
19 Kwon & Trail, 2005; Nagy, 2002). First, single-items have increased face validity in  
20 comparison to multi-item measures, particularly when the respondent deems the item  
21 representative of the construct as a whole (Nagy, 2002). Single-item measures are easier to  
22 interpret than their multi-item counterparts and if an individual understands the requirements  
23 of a measurement tool (i.e., what is being asked of them) they are more likely to provide a  
24 response. Second, single-items eliminate item redundancy and therefore reduce the fatigue,  
25 frustration, and boredom associated with answering similar questions repeatedly (Robins,

1 Hendin, & Trzesniewski, 2001). Finally, measures of this type lower the chance of common  
2 method variance (Gardner, Cummings, Dunham, & Pierce, 1998), where imitative  
3 correlations are observed due to the use of the same response format rather than the content  
4 of items (Williams, Cote, & Buckley, 1989).

5         Single-item measures hold advantages over multi-item measures in a number of  
6 practical circumstances. For example, when considering team performance in sport where  
7 time is constrained, single-item measures allow for in-game measurement. This is intuitively  
8 attractive, as group constructs such as collective efficacy comprise state beliefs that may vary  
9 during an event. Due to the short response time accompanying single-item measurement,  
10 they are ideally suited to the collection of quantitative data in research projects with large  
11 target populations (e.g., longitudinal studies; cf. Jordan & Turner, 2008). In addition, as the  
12 readiness of participants to complete questionnaires decreases as item numbers increase  
13 (Bean & Roszkowski, 1995; Edwards, Roberts, Sandercock, & Frost, 2004), single-item  
14 measures ensure maximal adherence of participants and prevent the likelihood of response set  
15 bias.

16         Single-item instruments also offer opportunities for the exploration of novel research  
17 domains beyond those currently available using traditional multiple-item scales. For  
18 example, studies have used individual items from existing multi-item psychometric scales as  
19 stimuli to trigger the brain activation associated with a specific psychological process (see  
20 e.g., Dimoka, 2011). Indeed, to comprehensively understand human processes and social  
21 constructs, such as collective efficacy, further integration of both brain and behavior  
22 assessment has been advocated (cf. Dimoka). In addition to the considerable literature  
23 examining behavioral and cognitive variables associated with collective efficacy (see Chow  
24 & Feltz, 2008 for a review in sport), the neuropsychological mechanisms that underpin  
25 individual perceptions of collective efficacy within groups have been considered (see Shearer

1 et al., 2009 for a full review). While studies have yet to directly examine the neural activity  
2 associated with the development of collective efficacy beliefs, neural correlates of other  
3 social psychological processes such as empathy (e.g., Carr, Iacoboni, Dubeau, Mazziotta, &  
4 Lenzi, 2003), and emotional recognition have been considered (e.g., Thom et al., 2012). The  
5 neuroscience literature shows that human processes activate many brain areas, and in turn,  
6 brain areas are activated by many different processes (Poldrack, 2006). A complex construct  
7 such as collective efficacy would therefore typically map onto more than one brain area,  
8 complicating the neural matching process.

9         Mapping of human processes to brain areas is accomplished by engaging participants  
10 in specific actions and observing their corresponding brain activations (Dimoka, 2011). To  
11 measure the neural correlates of collective efficacy it is necessary to stimulate individual  
12 perceptions over a short time period as this will heighten neural activity for the brain areas  
13 involved with collective efficacy development. A recent study has used observation-based  
14 interventions to enhance individual collective efficacy in team sports players in this manner  
15 (see Bruton, Mellalieu, & Shearer, 2014). In order to improve the measurement of  
16 psychological processes and their corresponding neural activity, it is suggested that  
17 psychometric scales need to be integrated with functional neuroimaging methods (cf.  
18 Dimoka). The instantaneous, simple response format associated with single-item instruments  
19 makes a single-item collective efficacy measure ideal for use with observation interventions  
20 and fMRI protocols. Combined, these two methods are likely to evoke a strong collective  
21 efficacy response and will allow for the collection of psychometric and brain mapping data  
22 simultaneously, permitting the accurate measurement of the neural activity associated with  
23 collective efficacy development.

24         According to Bandura (2006), all efficacy measures should be designed specifically  
25 for the intended context under examination. It is therefore unfeasible to create a single item

1 collective efficacy instrument that can be used in all contexts. To improve consistency  
2 between future studies, the alternative option is to employ a standardized operational stem  
3 that can be combined with any situational-specific context to form a single-item measure of  
4 collective efficacy. The current investigation reports the validation of an existing stem for  
5 use with single-item collective efficacy measurement across three studies using team sports  
6 participants. We chose the stem used with the CEQS (Short et al., 2005) '*Rate your team's*  
7 *confidence in their ability to...*' for this investigation. Study one examines the concurrent,  
8 convergent, and predictive validity of the stem incorporated in a single-item collective  
9 efficacy measure for use with competitive sports teams. Specifically, comparisons are made  
10 with validated full form measures of collective efficacy and team cohesion in a cross-  
11 sectional design with team sports participants. In study two the stem was included in a  
12 single-item measure for a laboratory-based task to manipulate the direction (positive, neutral,  
13 or negative) of collective efficacy beliefs with team sport participants. Concurrent and  
14 predictive validity of the measure was examined, along with test-retest reliability. Finally,  
15 study three assessed the validity and reliability of the stem (concurrent and predictive validity  
16 and test-retest reliability) using a field-based intervention design with interdependent team  
17 sports players.

### 18 **Study 1**

19 To test concurrent validity of the stem we examined the predictive capabilities of the CEQS  
20 towards the single-item measure. As the two instruments were hypothesized to measure the  
21 same construct, we suggested that both CEQS composite and subscale scores would predict  
22 single-item scores. To test the convergent validity of the stem we examined the predictive  
23 capabilities of the Group Environment Questionnaire (GEQ) towards the single-item  
24 measure. As collective efficacy and task cohesion exhibit a strong relationship (e.g., Kozub  
25 & McDonnell, 2000), we hypothesized that the two task components of the GEQ would

1 predict single-item collective efficacy scores. To test the predictive validity of the stem we  
2 examined the predictive capabilities of previous performance (win percentage over the  
3 previous three results) towards the single-item measure. As team performance is reported to  
4 predict collective efficacy (see Stajkovic et al., 2009, for a meta-analysis), we hypothesized  
5 that previous performance would predict single-item scores.

## 6 **Method**

### 7 **Participants**

8 311 interdependent sports team players ( $M_{\text{age}} = 21.69$ , age range: 16-54) were  
9 sampled from a variety of interdependent team sports in the United Kingdom and Ireland,  
10 including; football ( $n = 96$ ), rugby ( $n = 81$ ), hockey ( $n = 42$ ), netball ( $n = 28$ ), basketball ( $n =$   
11 20), cricket ( $n = 19$ ), lacrosse ( $n = 9$ ), badminton ( $n = 7$ ), Australian rules ( $n = 4$ ), American  
12 football ( $n = 4$ ), and handball ( $n = 1$ ). On average participants had played for their current  
13 team for 3.63 years ( $SD = 2.77$  years). The competitive level of the teams represented  
14 comprised collegiate ( $n = 244$ ), amateur ( $n = 61$ ), and semi-professional ( $n = 6$ ).

### 15 **Measures**

16 **Single-Item Measure of Collective Efficacy.** For the present study the stem was  
17 combined with a secondary component to form a single-item measure of collective efficacy  
18 for use with competitive sports teams: '*Rate your team's confidence in their ability to*  
19 *perform to a high level, in order to achieve success in their next competitive performance*'.  
20 The additional component of the item accounts for Bandura's (2006) recommendations that  
21 measures be phrased in terms of "can do" rather than "will do", treat efficacy beliefs as a  
22 state, and include information regarding the specific domain of functioning (i.e., competitive  
23 team sports performance). All responses to the single-item measure were rated on a  
24 confidence scale between 0 (not at all confident) and 100 (completely confident).



1           **Collective Efficacy Questionnaire for Sports (CEQS).** The CEQS (Short et al.,  
2 2005) was used as a validated measure of collective efficacy for comparison with the single-  
3 item measure. The CEQS is a 20-item questionnaire consisting of five factors (effort,  
4 persistence, ability, preparation, and unity) that can be combined to create a composite  
5 collective efficacy score. Ratings were made on a 10-point rating scale ranging from 0 (not  
6 at all confident) to 9 (completely confident). Construct validation of the measure with  
7 college-age student-athletes (Short et al.) using confirmatory factor analysis has indicated  
8 that the model is robust (CFI = .92, NNFI= .90, SRMR = .06), the exception being the error  
9 of approximation statistic (RMSEA = .10), which represents a mediocre fit (see Browne &  
10 Cudeck, 1993). Short et al. also reported strong internal reliability ( $\alpha$  range = .81 - .96), with  
11 similar findings evident for our study: Ability ( $\alpha$  = .93), Effort ( $\alpha$  = .83), Unity ( $\alpha$  = .80),  
12 Persistence ( $\alpha$  = .81), Preparation ( $\alpha$  = .85).

13           **The Group Environment Questionnaire.** The GEQ (GEQ; Carron, Widmeyer, &  
14 Brawley, 1985) was used to assess group cohesion for comparison with collective efficacy  
15 results for the single-item measure. The GEQ is an 18-item questionnaire consisting of four  
16 factors: Individual Attractions to the Group-Task (ATG-T), which reflects a member's  
17 feelings about their personal involvement with the group's task; Individual Attractions to the  
18 Group-Social (ATG-S), which reflects a member's feelings about their personal social  
19 interaction with the group; Group Integration-Task (GI-T), which reflects a member's  
20 perceptions of the similarity and unification of the group as a whole around their tasks and  
21 objectives; Group Integration-Social (GI-S), which reflects a member's perceptions of the  
22 similarity and unification of the group as a social unit. Responses are made on a 9-point  
23 likert scale between 1 (strongly disagree) and 9 (strongly agree). The original study reported  
24 acceptable internal reliability for each of the GEQ factors ( $\alpha$  range = .64 - .76) with mixed

1 findings evident for this study: ATG-T ( $\alpha = .60$ ), ATG-S ( $\alpha = .57$ ), GI-T ( $\alpha = .48$ ), GI-S ( $\alpha =$   
2  $.73$ ).

3 **Previous Performance.** As performance outcome (win/loss) predicts collective  
4 efficacy levels (Feltz & Lirgg, 1988), participants recorded their team's performance record  
5 for their three most recent competitive performances to form a win percentage (cf. Bruton,  
6 Mellalieu, Shearer, Roderique-Davies, & Hall, 2013).

### 7 **Procedure**

8 Ethical approval was granted by a university ethics committee for all three studies,  
9 and all participants provided informed consent before taking part. Prior to the beginning of  
10 the competitive season the research team created a questionnaire pack that included a  
11 demographic sheet, the single-item measure, the CEQS, and the GEQ. During the  
12 competitive season interdependent sports team players were provided with a link to an online  
13 version of the questionnaire pack, developed using an online-survey provider  
14 ([www.surveymonkey.com](http://www.surveymonkey.com)). Prior to participation individuals were informed that their  
15 involvement in the study was voluntary, there was no correct/incorrect answer to any of the  
16 questions provided, and that answers would remain strictly confidential and securely stored  
17 on computers within the university department of the research team. The online  
18 questionnaire pack took approximately ten minutes to complete.

### 19 **Data Analysis**

20 All statistical procedures for the studies were conducted using a minimum  
21 significance level of  $p = 0.05$ . First, data were screened for univariate normality, multivariate  
22 normality, and multicollinearity. Second, using Miller, Meier, Muehlenkamp, and  
23 Weatherly's (2009) recommendations, regression analyses were used to examine validity.  
24 Specifically, two simple regression analyses were used to examine whether composite CEQS  
25 score and previous performance (win %) predicted single-item collective efficacy

1 respectively. Similarly, two forced entry multiple regression analysis were used respectively  
2 to examine whether CEQS subscales and GEQ dimensions were predictive of collective  
3 efficacy measured using the single-item (i.e., the direction and relative contribution of each  
4 variable towards the variance in collective efficacy scores).

## 5 **Results**

### 6 **Data Screening**

7 Cook's distances were used to examine the assumptions of multivariate normality,  
8 with a value greater than 1 indicative of multivariate outliers (cf. Cook & Weisberg, 1982).  
9 For all regression analyses Cook's distance values were below 1 with a maximum value of  
10 0.24, indicating that no single case had a large influence on the respective model, leaving 311  
11 cases for each analysis. The VIF values were all below 10 with an average close to 1, and the  
12 tolerance statistics were above 0.2, indicating no collinearity within the data.

### 13 **Concurrent Validity**

14 The relationship between collective efficacy responses to the CEQS and the single-  
15 item measure were assessed using two regression analyses. The first simple regression  
16 analysis identified that the composite CEQS score accounted for 48% of variability in  
17 collective efficacy using the single-item measure ( $\beta = .69$ ,  $R^2$  change = .48,  $F_{1-309} = 280.28$ ,  $p$   
18  $< 0.001$ ). The second multiple regression analysis reported that CEQS subscales accounted  
19 for 56% of variability in collective efficacy measured using the single-item. The analysis  
20 identified that Ability ( $\beta = .51$ ,  $R^2$  change = .56,  $F_{5-305} = 76.14$ ,  $p < 0.001$ ) and Persistence ( $\beta$   
21  $= .15$ ,  $R^2$  change = .56,  $F_{5-305} = 76.14$ ,  $p < 0.05$ ) subscales of the CEQS were the only  
22 significant predictors of collective efficacy using the single-item.

### 23 **Convergent Validity**

24 The relationship between group cohesion and collective efficacy was assessed using a  
25 multiple regression analysis. The regression analysis reported that the GEQ dimensions

1 accounted for 17% of variability in collective efficacy. ATG-T ( $\beta = .18$ ,  $R^2$  change = .17,  $F_{4-306} = 15.15$ ,  $p < 0.01$ ), GI-T ( $\beta = .22$ ,  $R^2$  change = .17,  $F_{4-306} = 15.88$ ,  $p < 0.01$ ), and GI-S ( $\beta =$   
2  $.16$ ,  $R^2$  change = .17,  $F_{4-306} = 15.88$ ,  $p < 0.05$ ) dimensions were identified as significant  
3 predictors towards single-item collective efficacy scores. The ATG-S dimension was not a  
4 significant predictor of the single-item score ( $p > 0.05$ ).

### 6 **Predictive Validity**

7 The relationship between previous performance and collective efficacy was assessed using a  
8 simple regression analysis. The regression analysis reported that 17% of variability in  
9 collective efficacy was accounted for by Previous Performance scores ( $\beta = .41$ ,  $R^2$  change =  
10  $.17$ ,  $F_{1-309} = 63.84$ ,  $p < 0.001$ ).

## 11 **Discussion**

12 In study one, concurrent validity for the operational stem was supported with  
13 composite CEQS scores identified as a significant predictor of collective efficacy measured  
14 using the single-item. The findings also supported the convergent validity for the stem, with  
15 the ATG-T, GI-T, and GI-S components of the GEQ identified as significant predictors of  
16 single-item collective efficacy. Finally, the predictive validity of the stem was demonstrated  
17 with previous performance predicting single-item collective efficacy scores.

## 18 **Study 2**

19 To further test the concurrent and predictive validity of the stem we examined the  
20 relationship between the CEQS and the single-item measure before and after an intervention  
21 designed to manipulate collective efficacy. Recent research has shown observation  
22 interventions can be used to manipulate collective efficacy beliefs in teams (Bruton et al.,  
23 2014), therefore, group-specific observation interventions (positive/neutral/negative) were  
24 tailored to manipulate collective efficacy beliefs. The single-item measure was predicted to  
25 distinguish collective efficacy scores according to the expected direction of the intervention

1 effects. Specifically, individuals allocated to a negative observation condition would  
2 experience decreased efficacy, allocation to the neutral condition would result in no change,  
3 and allocation to a positive condition would increase efficacy. Finally, test-retest reliability  
4 of the stem was investigated using pre- and post-intervention single-item scores for two  
5 subsamples. Based on the procedures adopted when examining test-retest reliability for a  
6 previously developed anxiety measure (cf. Williams, Morlock, & Feltner, 2010), the first  
7 sample included participants that experienced little/no change in CEQS scores ( $\leq 0.2$ )  
8 between pre- and post-intervention measures. As the simple regression from study one  
9 showed CEQS scores predicted single-item scores, participants that exhibit little/no change in  
10 CEQS scores were predicted to show little/no change in single-item scores. Based on the  
11 assumption that the neutral intervention would have no effect on collective efficacy  
12 perceptions, the second sample included participants allocated to the neutral intervention  
13 condition. A positive correlation was predicted to exist between pre- and post-intervention  
14 collective efficacy scores for the single-item measure for both samples.

## 15 **Method**

### 16 **Participants**

17 One hundred and thirty three undergraduate students ( $M_{age} = 20.63$  years,  $SD_{age} = 1.84$   
18 years) from a higher education institution in South Wales, UK were recruited via opportunity  
19 sampling. Each participant played in an interdependent team sport, including; rugby union ( $n$   
20  $= 42$ ), soccer ( $n = 36$ ), field hockey ( $n = 24$ ), basketball ( $n = 18$ ), and netball ( $n = 13$ ),  
21 ensuring a degree of familiarity with teamwork, physical activity, and group dynamics (i.e.,  
22 collective beliefs).

### 23 **Measures**

24 **Single-Item Measure of Collective Efficacy.** In this study the stem was combined  
25 with a secondary component to form a single-item measure of collective efficacy for use with

1 the experimental task: *'Rate your team's confidence in their ability to complete the obstacle*  
2 *course in the shortest possible time'*. In line with the procedure of study one the second  
3 component of the single-item measure was constructed to account for Bandura's (2006)  
4 guidelines for the development of efficacy scales. All responses to the single-item measure  
5 were rated on a confidence scale between 0 (not at all confident) and 100 (completely  
6 confident) using an adaptive visual analogue scale (see Marsh-Richard, Hatzis, Mathias,  
7 Venditti, & Dougherty, 2009 for a full description of this measurement type).

8 **Collective Efficacy Questionnaire for Sports.** As in study one, the CEQS (Short et  
9 al., 2005) was employed as a criterion measure for individual-level perceptions of collective  
10 efficacy. Cronbach alpha coefficients indicated adequate internal reliability for the sample:  
11 Ability ( $\alpha = .89$ ), Effort ( $\alpha = .84$ ), Unity ( $\alpha = .81$ ), Persistence ( $\alpha = .83$ ), and Preparation ( $\alpha =$   
12  $.80$ ).

### 13 **Procedure**

14 The experiment was fifteen-days in duration with participants required to attend the  
15 laboratory on day one and day fifteen respectively. To maximize motivation participants  
16 were told that they were to participate in a UK-wide experiment on teamwork, competing in a  
17 complex task requiring balance, co-ordination, and communication, characteristics desirable  
18 in the performance of sports teams. To ensure the task was competitive teams were led to  
19 believe that they were participating as representatives of their university against other  
20 university teams. This was established by showing the participants a false datasheet, with a  
21 large sample size and names of UK-wide universities.

22 A single-blinded randomized design was adopted identical to that used in a recent  
23 collective efficacy investigation (see Bruton et al., 2014 for detailed information). Members  
24 of the same sports team were placed into teams of three and randomly and blindly allocated  
25 to one of three treatment groups (i.e., positive,  $n = 16$ ; neutral,  $n = 14$ ; or negative,  $n = 15$ ).

1 Once assembled each team participated in three practice trials for an obstacle-based task  
2 (Figure 1), after which they were provided with a false average performance time lying in the  
3 middle tenth of a fictitious database across other UK universities. Participants were asked to  
4 return to the laboratory in fourteen days time and informed that they would participate in a  
5 competitive trial to be used for the UK wide experiment. All practice sessions were video  
6 recorded for the purpose of developing team-specific video interventions during this  
7 fourteen-day break period. The interventions were condition-based, meaning groups  
8 allocated to the positive condition viewed positive video clips of their team (i.e., footage  
9 displaying five examples of successful team performance for obstacles used in the task) and  
10 groups allocated to the negative condition viewed negative video clips of their team (i.e.,  
11 footage displaying five examples of unsuccessful team performance for obstacles used in the  
12 task) collected from their respective practice performances. For the neutral condition a  
13 standardized video intervention was adopted displaying footage depicting the layout of the  
14 obstacle course used in the experimental task.

15 When participants returned to the laboratory on day 15, they were reminded of both  
16 the task requirements and their mediocre results in the practice trials. Each of the teams  
17 completed the CEQS and the single-item measure for the first time (pre-intervention), after  
18 which they were informed they would take part in the competitive trial in thirty minutes.  
19 Upon completion of this first measure, their respective intervention strategies were  
20 administered. Once the intervention was complete, collective efficacy responses were  
21 recorded for the second time (post-intervention) using the CEQS and single-item measure  
22 and the participants debriefed about the real purpose of the experiment.

### 23 **Data Analysis**

24 Data was screened for normality and homogeneity of variance using the Shapiro-Wilk  
25 test and Levene's test respectively. A bivariate, one-tailed Pearson's correlation was used to

1 examine the relationship between the collective efficacy scores for the single-item measure  
2 and the composite and subscale scores for the CEQS. In addition, confidence intervals were  
3 computed for all of the correlations. A mixed 3 x 2 (*condition* x *time*) model analysis of  
4 variance (ANOVA) was used to examine the predictive validity of the single-item collective  
5 efficacy scores for main effects and interactions of the independent variables. Specifically,  
6 *condition* (*positive/neutral/negative*) was used as the between-subjects factor, while *time*  
7 (*pre-intervention/post-intervention*) was used as the within-subjects factor. Simple planned  
8 contrasts were used to make comparisons between time (reference category: pre-intervention)  
9 and condition (reference category: negative). Test-retest reliabilities for the single-item  
10 scores were computed for the two aforementioned subsamples. Intraclass correlation  
11 coefficients (ICCs) were computed using a two-way (*participants* x *time*) random effects  
12 ANOVA as recommended by Williams et al. (2010) in their design examining the validity of  
13 a single-item anxiety measure.

## 14 Results

### 15 Data Screening

16 Collective efficacy data for each group was screened for assumptions of normality  
17 both pre- and post-intervention. The Shapiro-Wilk test identified CEQS and single-item  
18 measure scores for the positive ( $D(48) = .95-.99, p > .05$ ), neutral ( $D(41) = .94-.97, p > .05$ ),  
19 and negative groups ( $D(44) = .90-.98, p > .05$ ), as normal at both time points. The Levene's  
20 test reported equal variance in CEQS and single-item measure scores for all conditions both  
21 pre- ( $F(2, 130) = 0.38, 2.45, p > .05$ ) and post-intervention ( $F(2, 130) = 5.20, 7.43, p > .05$ ).

### 22 Concurrent Validity

23 The single-item measure reported significant correlations with the criterion measure  
24 for collective efficacy. Specifically, the single-item scores were highly correlated with the  
25 composite CEQS scores both pre- ( $r = .48, p < .001, 95\% \text{ CI } [.34, .60]$ ) and post-intervention



1 ( $r = .73, p < .001, 95\% \text{ CI } [.64, .80]$ ). The single-item also correlated strongly with each of  
2 the five subscales for the CEQS at both time points. At pre-intervention the strongest  
3 correlation was reported for Ability ( $r = .50, p < .001, 95\% \text{ CI } [.36, .62]$ ), with additional  
4 positive correlations for Preparation ( $r = .39, p < .001, 95\% \text{ CI } [.24, .53]$ ), Persistence ( $r = .35,$   
5  $p < .001, 95\% \text{ CI } [.19, .49]$ ), Unity ( $r = .34, p < .001, 95\% \text{ CI } [.18, .48]$ ), and Effort ( $r = .28, p <$   
6  $.001, 95\% \text{ CI } [.12, .43]$ ). Post-intervention Ability showed the strongest correlation ( $r = .71, p$   
7  $< .001, 95\% \text{ CI } [.62, .79]$ ), with additional positive correlations for Preparation ( $r = .66, p <$   
8  $.001, 95\% \text{ CI } [.55, .75]$ ), Persistence ( $r = .63, p < .001, 95\% \text{ CI } [.52, .73]$ ), Unity ( $r = .62, p <$   
9  $.001, 95\% \text{ CI } [.51, .72]$ ), and Effort ( $r = .63, p < .001, 95\% \text{ CI } [.52, .73]$ ).

## 10 **Predictive Validity**

11 The mixed 3 x 2 ANOVA results for the single-item collective efficacy scores  
12 suggested a significant main effect within groups for time, between pre- and post-intervention  
13 measures ( $F(1,130) = 75.96, p < .05, r = .61$ ), a significant main effect between groups for  
14 condition ( $F(2,130) = 32.57, p < .05, r = .45$ ) and a significant interaction between time and  
15 condition ( $F(2,130) = 43.97, p < .05, r = .50$ ). Inspection of the score profiles indicated the  
16 nature of the difference between the three conditions (Figure 2). Pre-intervention collective  
17 efficacy scores (Table 1) indicated little difference between the positive ( $M = 73.15, SD =$   
18  $7.85$ ), neutral ( $M = 69.29, SD = 11.75$ ) and negative conditions ( $M = 66.64, SD = 14.05$ ).  
19 Simple planned contrasts showed post-intervention differences between the positive and  
20 neutral conditions ( $M_{diff} = 5.81, SE = 2.34, p < .05$ ), positive and negative conditions ( $M_{diff} =$   
21  $18.28, SE = 2.30, p < .05$ ), and the neutral and negative conditions ( $M_{diff} = 12.47, SE = 2.39,$   
22  $p < .05$ ). A decrease was observed in mean scores for the positive ( $M = 72.65, SD = 9.80$ ),  
23 neutral ( $M = 64.89, SD = 13.61$ ), and negative conditions ( $M = 42.60, SD = 17.59$ ).

24

25

## 1 **Test-Retest Reliability**

2           The two-way random effects ANOVA results for the single-item collective efficacy  
3 scores suggested adequate test-retest reliability over time. For the individuals that showed  
4 the smallest change ( $\leq .2$ ) in CEQS scores, a large single measure ICC ( $r = .77, p < .05, CI =$   
5  $.47-.90$ ) was reported between pre- and post-intervention measures ( $M_{diff} = 4.94$ ). For the  
6 individual's allocated to the neutral intervention condition, a moderate single measure ICC ( $r$   
7  $= .62, p < .05, CI = .38-.79$ ) was reported between pre- and post-intervention measures ( $M_{diff}$   
8  $= 4.40$ ).

## 9 **Discussion**

10           Study two reported strong correlations between the single-item measure and CEQS  
11 for interdependent team sports players in a lab-based experiment, further substantiating the  
12 concurrent validity of the stem. The results also provide support for the stem's predictive  
13 validity. Although single-item collective efficacy scores decreased post-intervention for all  
14 three conditions, the largest decrease coincided with the negative condition, and the smallest  
15 decrease existed for the positive condition. Whilst it is reasonable to expect an increase for  
16 the positive condition, no change for the neutral condition, and a decrease for the negative  
17 condition, this was not apparent for the CEQS. CEQS scores increased for the positive and  
18 neutral conditions, and decreased for the negative group. However, post-intervention  
19 collective efficacy beliefs were highest for the positive condition and lowest for the negative  
20 condition using both measures. Lastly, the findings indicated test-retest reliability for the  
21 stem as strong correlations were reported between pre-and post-intervention single-item  
22 scores using participants allocated to the neutral observation condition and individuals that  
23 demonstrated little/no difference in corresponding CEQS scores.

24

25

### 1 Study 3

2 To examine the concurrent validity of the stem we predicted that CEQS composite  
3 and subscales scores would hold a positive correlation with single-item scores based on the  
4 relationship demonstrated between the two measures in studies one and two. To test the  
5 predictive validity of the stem we measured collective efficacy levels before and after an  
6 intervention. Previous studies using modeling techniques that involve the observation of  
7 oneself or others, have shown its influence on both psychological processes and task  
8 performance to be greatest when similarity between the model and the observer is high (e.g.,  
9 Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005, 2006; Short & Ross-Stewart,  
10 2009). Consequently, observing positive footage of one's own team performing familiar  
11 activities should increase collective efficacy, whilst observing positive footage of an  
12 unknown team performing unfamiliar activities should have little impact upon efficacy  
13 beliefs. The single-item measure was predicted to distinguish collective efficacy scores  
14 according to the expected direction of the intervention effects. When viewing positive  
15 basketball footage, members of a basketball team (sample A) would experience increased  
16 efficacy, and members of other interdependent sports teams (sample B) would experience no  
17 change in efficacy perceptions. Finally, test-retest reliability of the stem was examined in the  
18 same manner as study two, using pre- and post-intervention single-item scores for two  
19 subsamples. The first sample included participants that experienced little/no change in CEQS  
20 scores between pre- and post-intervention measures (cf. Williams, Cote, & Buckley, 2010),  
21 and the second sample included participants who were allocated to the unfamiliar observation  
22 intervention condition. It was predicted that a positive correlation would exist between pre-  
23 and post-intervention collective efficacy scores for the single-item measure using both  
24 samples.

25

## 1 **Method**

### 2 **Participants**

3 Male participants ( $N = 36$ ) were recruited via opportunity sampling from a university  
4 basketball squad ( $n = 18$ ,  $M_{\text{age}} = 21.73$  years,  $SD_{\text{age}} = 1.51$  years) and other interdependent  
5 sports teams ( $n = 18$ ,  $M_{\text{age}} = 21.94$  years,  $SD_{\text{age}} = 1.76$  years) from a South Wales university.  
6 The basketball players competed for either the men's 1<sup>st</sup> team or 2<sup>nd</sup> team in British  
7 Universities and Colleges competition. They were recruited because the controlled  
8 environment for competitive fixtures allowed for the collection of detailed video footage for  
9 use with the observation interventions. Interdependent team sports players were recruited  
10 from other team sports at the same institution (rugby union, soccer, and field hockey) because  
11 of their understanding of competitive sport, and their relative lack of understanding of  
12 basketball performance (i.e., no competitive experience). Together, these two sub-samples  
13 provided an opportunity to examine the effect of content familiarity upon collective efficacy  
14 responses to positively oriented video footage of competitive basketball.

### 15 **Measures**

16 **Single-Item Measure of Collective Efficacy.** In this design the stem was included in  
17 a single-item measure identical to that used in study one, namely '*Rate your team's*  
18 *confidence in their ability to perform to a high level, in order to achieve success in their next*  
19 *competitive performance*'. All responses were rated on a confidence scale between 0 (not at  
20 all confident) and 100 (completely confident).

21 **Collective Efficacy Questionnaire for Sports.** The CEQS (Short, Sullivan, & Feltz,  
22 2005) was employed as a criterion measure for individual-level perceptions of collective  
23 efficacy. Cronbach alpha coefficients indicated adequate internal reliability for the sample:  
24 Ability ( $\alpha = .88$ ), Effort ( $\alpha = .82$ ), Unity ( $\alpha = .81$ ), Persistence ( $\alpha = .81$ ), Preparation ( $\alpha =$   
25  $.74$ ).

## 1 **Procedure**

2           Video footage of seventeen competitive fixtures was collected for two university  
3 basketball teams over an 8-week period. Footage consisted of actual performance (on court),  
4 team interactions during performance (i.e., communication, team drills), and reactions to  
5 performance results (both on and off court, i.e., successful baskets/plays). Recordings  
6 focused on positive video footage (i.e., a celebratory reaction to success, a performer being  
7 pleased with performance, a successful completion of an action, a significant performance  
8 result). The team-specific observation interventions were tailored to include each team  
9 member in at least two of the video clips, and involve all aspects of overall basketball  
10 performance. Subsequently, a ninety second video compiling seven separate video clips was  
11 developed for each of the basketball teams' *familiar* observation intervention, and the non-  
12 basketball participants were randomly allocated either the 1<sup>st</sup> ( $n = 9$ ) or 2<sup>nd</sup> ( $n = 9$ ) basketball  
13 team intervention for their *unfamiliar* observation intervention.

14           Data collection comprised a three-step process. First, participants completed the  
15 CEQS and single-item measure (pre-intervention), after which the intervention was  
16 administered. Once the observation intervention was watched in full, collective efficacy  
17 beliefs were once again collected using both measures (post-intervention) and participants  
18 were debriefed on the purpose of the study.

## 19 **Data Analysis**

20           Data was screened for normality and homogeneity of variance using the Shapiro-Wilk  
21 test and Levene's test respectively. A bivariate, one-tailed Pearson's correlation was used to  
22 examine the relationship between the single-item collective efficacy scores and CEQS scores  
23 (composite & subscales). In addition, confidence intervals were computed for all of the  
24 correlations. A mixed 2 x 2 (*familiarity x time*) model ANOVA was used to examine the  
25 data for main effects and interactions of the independent variables for the single-item scores.

1 Specifically, *familiarity (familiar/unfamiliar)* was used as the between-subjects factor, while  
2 *time (pre-intervention/post-intervention)* was used as the within-subjects factor. Test-retest  
3 reliabilities for the single-item measure were computed using the two aforementioned  
4 subsamples. Specifically, ICCs were computed using a two-way (*participants x time*)  
5 random effects ANOVA. All statistical procedures were conducted using SPSS for  
6 Windows, version 20, utilizing a minimum significance level of  $p = 0.05$ .

## 7 Results

### 8 Data Screening

9 Collective efficacy data for each group was screened for the assumptions of normality  
10 at both pre- and post-intervention. The Shapiro-Wilk test indicated that CEQS and single-  
11 item data for the familiar ( $D(18) = .87-.92, p > .05$ ) and unfamiliar groups ( $D(18) = .94-.95,$   
12  $p > .05$ ) was normal at both time points. The Levene's test reported equal variance in  
13 collective efficacy scores for both groups pre- ( $F(1, 34) = .49-.51, p > .05$ ) and post-  
14 intervention ( $F(1, 34) = .01-2.2, p > .05$ ).

### 15 Concurrent Validity

16 The single-item measure scores were strongly correlated with the composite CEQS  
17 scores both pre- ( $r = .74, p < .001, 95\% \text{ CI } [.54, .86]$ ) and post-intervention ( $r = .69, p < .001,$   
18  $95\% \text{ CI } [.47, .83]$ ). The single-item measure also showed positive correlations with each of  
19 the five subscales for the CEQS. Pre-intervention the single-item correlated strongly with  
20 Ability ( $r = .67, p < .001, 95\% \text{ CI } [.44, .82]$ ), Preparation ( $r = .64, p < .001, 95\% \text{ CI } [.39, .80]$ ),  
21 Persistence ( $r = .66, p < .001, 95\% \text{ CI } [.42, .81]$ ), Unity ( $r = .54, p < .001, 95\% \text{ CI } [.26, .74]$ ),  
22 and Effort ( $r = .62, p < .001, 95\% \text{ CI } [.37, .79]$ ). Post-intervention the single-item correlated  
23 positively with Ability ( $r = .57, p < .001, 95\% \text{ CI } [.30, .76]$ ), Preparation ( $r = .66, p < .001, 95\%$   
24  $\text{CI } [.42, .81]$ ), Persistence ( $r = .68, p < .001, 95\% \text{ CI } [.45, .82]$ ), Unity ( $r = .38, p < .05, 95\% \text{ CI}$   
25  $[.06, .63]$ ), and Effort ( $r = .68, p < .001, 95\% \text{ CI } [.45, .82]$ ).

## 1 **Predictive Validity**

2           The mixed 2 x 2 ANOVA results for the single-item collective efficacy scores  
3 suggested a significant main effect within groups for time, between the pre- and post-  
4 intervention measures ( $F(1,34) = 33.66, p < .05, r = .66$ ), no effect between groups for  
5 familiarity ( $F(1,34) = .37, p > .05, r = .10$ ) and a significant interaction between time and  
6 familiarity ( $F(1, 34) = 21.84, p < .05, r = .62$ ). Closer inspection of the score profiles  
7 indicated the nature of the difference between the groups (Figure 3). Pre-intervention  
8 collective efficacy scores (Table 2) identified that the familiar group ( $M = 74.61, SD = 9.91$ )  
9 had a lower mean score than the unfamiliar group ( $M = 76.94, SD = 8.34$ ). However, an  
10 increase was observed in post-intervention mean scores for both the familiar ( $M = 83.56, SD$   
11  $= 6.26$ ) and unfamiliar group ( $M = 77.94, SD = 8.83$ ), with the increase greatest for the  
12 familiar group.

## 13 **Test-Retest Reliability**

14           The two-way random effects ANOVA results for the single-item measure suggested  
15 strong test-retest reliability over time for both samples. For individuals who demonstrated  
16 little/no change in CEQS scores, a large single measure ICC ( $r = .88, p < .05, CI = .44-.98$ )  
17 was reported between pre- and post-intervention single-item scores ( $M_{diff} = 0.43$ ). For  
18 individuals allocated to the unfamiliar observation intervention, there was a large single  
19 measure ICC ( $r = .87, p < .05, CI = .70-.95$ ) between pre- and post-intervention single-item  
20 scores ( $M_{diff} = 0.89$ ).

## 21 **Discussion**

22           The findings from study three replicated the correlations between the single-item and  
23 CEQS from the first two studies in a field-based setting, further demonstrating the concurrent  
24 validity of the stem. A greater increase in collective efficacy was reported for the group  
25 familiar with the content of the observation-intervention in comparison to the unfamiliar

1 group, indicating predictive validity for the single-item measure. Finally, the test-retest  
2 reliability for the stem was supported with strong correlations between pre-and post-  
3 intervention single-item scores using participants allocated to the unfamiliar observation  
4 condition and individuals that demonstrated little/no difference in corresponding CEQS  
5 scores.

### 6 **General Discussion**

7 Taken together our findings support the validity and reliability of the operational stem for use  
8 with single-item collective efficacy measurement with team sports athletes. Correlations  
9 between the single-item measures and the CEQS were consistently strong, especially when  
10 compared to existing single-item measures of self-efficacy and their multi-item counterparts  
11 (e.g., Hoepfner, Kelly, Urbanoski, & Slaymaker, 2011,  $r = .30-.56$ ). The high concurrence  
12 with the CEQS and each of its subscales was consistent across the three different study  
13 designs (cross-sectional/laboratory/field) using team sports participants, demonstrating the  
14 rigor of the operational stem for collective efficacy measurement in a sports domain. The  
15 results of the investigation also indicated that the ability dimension of the CEQS was the  
16 most significant predictor of single-item collective efficacy for study one, and the most  
17 correlated dimension with single-item collective efficacy for study two (pre- and post-  
18 intervention) and three (pre-intervention). These findings may have occurred because the  
19 word 'ability' is used in the single-item stem and the items used for this subscale were  
20 worded specifically to reference the ability of the team, otherwise known as a 'wording effect'  
21 (cf. Horan, DiStefano, & Motl, 2009). The current investigation also considered the issue of  
22 reliability when examining the psychometric properties of the stem. Across study two and  
23 three high test-retest reliabilities were evident for four different subgroups using two different  
24 study designs, supporting the reliability of the stem for use with single-item collective  
25 efficacy measurement with team sports participants.



1           In this investigation, we have shown that the stem can be used as part of two different  
2 single-item measures and employed across three different study designs in a sports context.  
3 Due to its adaptability, the stem allows for greater consistency when measuring collective  
4 efficacy using single-item scales, a characteristic which Bandura (1997) deems necessary for  
5 the measurement of collective efficacy in future research. Whilst allowing for greater  
6 consistency, this approach also takes into account Bandura's (2006) guidelines for the  
7 construction of efficacy scales. Bandura recommends that efficacy measures are context-  
8 specific, treat efficacy beliefs as a state, and are phrased in terms of 'can do' rather than 'will  
9 do'. We suggest this information be included as a second component combined with the  
10 stem to form a single-item measure. While two single-item collective efficacy scales have  
11 been used in previous collective efficacy research (Greenlees, Graydon, & Maynard, 1999;  
12 Spink, 1990), the psychometric properties of these measures were not examined. Our results  
13 suggest the stem is a valid and reliable method for use with single-item collective efficacy  
14 measurement in sport. This scale type allows for the instantaneous measurement of  
15 collective efficacy, something that is beneficial considering the construct is a state belief (i.e.,  
16 can vary based on situational factors). This approach allows researchers and practitioners to  
17 examine collective efficacy at any given moment, providing a flexible measurement tool that  
18 can be used for both repeated measures and case study designs, promoting multi-level  
19 research within groups as well as between different contexts. For example, the stem could be  
20 used to measure collective efficacy at various time points during a sports team's training  
21 session to examine the relative impact of different coaching strategies or leadership tactics  
22 upon collective efficacy beliefs.

23           In the past decade experimental psychologists have identified the need for integration  
24 of brain and behavior assessment to gain a greater understanding of human psychological  
25 processes (cf. Henson, 2005). In this respect, a major reason for this investigation was to

1 develop a single-item measure that can be used as part of a neuroimaging study to examine  
2 the neuropsychological processes involved with collective efficacy. The majority of fMRI  
3 studies investigating the neural correlates of psychological processes have involved  
4 procedures designed to evoke a desired psychological response. For example, when  
5 exploring the brain activity linked to empathy, Rameson, Morelli, and Lieberman (2011)  
6 showed participants sentences and images depicting sad situations. However, research has  
7 advocated that psychometric scales be integrated within fMRI designs to assess  
8 brain/behavior more accurately. Specifically, Dimoka (2011) used items from multi-item  
9 psychometric scales for four psychological processes (trust, distrust, perceived usefulness,  
10 perceived ease of use) as a means to stimulate brain activity associated with each focal  
11 psychological process. We suggest the stem developed in the current study is suitable for use  
12 with single-item collective efficacy measurement in fMRI protocols. Moreover, the single-  
13 item scale can be combined with observation interventions designed to increase efficacy  
14 beliefs as a comprehensive method for assessing the brain activity associated with collective  
15 efficacy (see e.g., Bruton et al., 2014 for group-based observation interventions). Functional  
16 neuroimaging data provides an additional dependent variable that can be combined with  
17 behavioral data to inform psychological theory and further understand a psychological  
18 process (Henson, 2005). Successful mapping of the brain activity associated with collective  
19 efficacy will advance understanding of this construct, providing information that can be used  
20 in the refinement of existing theories, development of conceptually grounded intervention  
21 techniques, and production of assessment methods combining brain, behavior, and  
22 psychometric modes.

23         Although we believe the stem is both valid and reliable in the context of our  
24 investigation, there are some limitations that need to be acknowledged. Specifically, there is  
25 a need to consider the utility of single-item measures, the group cohesion measure used for

1 study one, the small population used for study three, and the small subsamples used for the  
2 reliability analyses in study two and three respectively. First, it is important to note that the  
3 collective efficacy literature has predominantly used multi-item instruments that consider  
4 several factors contributing towards a team's overall efficacy beliefs (see e.g., CEQS:  
5 Ability, Effort, Persistence, Preparation, Unity) and can therefore provide insights into the  
6 dynamics of team behavior (Bandura, 2006). Knowing that a group is confident or not is of  
7 upmost importance, but in order to aid the development of a group further their confidence in  
8 specific group processes needs to be considered. Using a single-item tool to measure  
9 collective efficacy does not inform the researcher/practitioner about an individual's efficacy  
10 perceptions for specific aspects of group performance. Single-item instruments should only  
11 be used to measure collective efficacy when multi-item tools are unfeasible (i.e., situations  
12 that accommodate little response time) or not warranted (i.e., only interested in overall  
13 collective efficacy response).

14 In study one, the GEQ was used as a multidimensional measure of group cohesion to  
15 examine the convergent validity of the stem. Previous research shows task components of  
16 group cohesion are related to collective efficacy (e.g., Kozub & McDonnell, 2000). Indeed,  
17 of all the GEQ subscales Short et al. (2005) suggest that GI-T should hold the strongest  
18 relationship with collective efficacy because this subscale considers the group and the task.  
19 Our findings showed GI-T was not a significant predictor, however, the ATG-T dimension  
20 significantly predicted collective efficacy. This suggests that liking your team's style of play  
21 relates to your view of your team's confidence. Although the GEQ is a valid and reliable  
22 measure for use in sport (e.g., Whitton & Fletcher, 2014), the ATG-S and GI-T components  
23 did show poor reliability in this instance. We suggest that future studies further examine the  
24 stem's validity using the GEQ with team sports players.

1           In study three the development of the intervention required collection of video  
2 footage for both basketball teams across several fixtures. This meant it was only possible to  
3 use two basketball teams and subsequently our study population couldn't exceed thirty-six as  
4 to avoid any biases. Despite this small population size, the within-subject and interaction  
5 effect sizes for this study ( $>.50$ ) are classified as a large effect within previous guidelines  
6 (Cohen, 1992), supporting the strength of the observation effect and the predictive validity of  
7 the stem. In both study two and three the reliability of the measure was considered across  
8 two different study designs. In the absence of the internal reliability statistic due to the  
9 single-item nature of the measure, we must consider its reliability over time. Because we  
10 used observation-based interventions in both studies, the number of participants predicted to  
11 show no change in collective efficacy was comparatively small to that used to measure  
12 reliability in previous studies (e.g., Williams et al., 2010). However, the findings are similar  
13 to the reliability statistics reported for other domains and outcomes, such as pain (Chang,  
14 Hwang, & Feuerman, 2000) and anxiety (Williams et al.). There is a need to further examine  
15 the stem using groups/teams with large populations or multiple groups/teams in order to  
16 ensure that desired observation effects are attainable. In this investigation we assessed  
17 collective efficacy at specific time points (e.g., before and after an intervention for studies  
18 two and three). Although we attempted to control for change in collective efficacy, the  
19 reliability of our single-item measure across several time points also requires consideration.  
20 A similar design to that adopted in study three can therefore be used to record collective  
21 efficacy responses for team sports players viewing an unfamiliar observation intervention  
22 over an extended period of time (i.e., repetitive viewing over several weeks).

23           In summary, our findings provide preliminary evidence for the use of this stem in  
24 future single-item collective efficacy measures. However, there is a need to further examine  
25 the psychometric properties of the stem for collective efficacy measurement in sports

1 contexts. To provide a more thorough assessment of the stem, research should focus on the  
2 relations between collective efficacy and other group related constructs, such as  
3 communication and group member satisfaction. With specific reference to sport, invariance  
4 tests for age, gender, level of competition, type of sport, and level of sport are required to  
5 fully validate this measure for use in this context (cf. Short et al., 2005), with multiple sports  
6 teams, research designs, and settings (i.e., in both training and competition).

7

## References

- 1  
2 Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*,  
3 2, 122-147.
- 4 Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman and  
5 Company.
- 6 Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan  
7 (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich, CT: Information  
8 Age Publishing.
- 9 Bean, A. G., & Roszkowski, M. J. (1995). The long and short of it. *Marketing Research, 7, 1*,  
10 20-26.
- 11 Bruton, A. M., Mellalieu, S. D., & Shearer, D. A. (2014). Observation interventions as a  
12 means to manipulate collective efficacy in groups. *Journal of Sport and Exercise*  
13 *Psychology, 36*, 27-39.
- 14 Bruton, A. M., Mellalieu, S. D., Shearer, D., Roderique-Davies, G., & Hall, R. (2013).  
15 Performance accomplishment information as predictors of self-efficacy as a function of  
16 skill level in amateur golf. *Journal of Applied Sport Psychology, 25, 2*, 197-208.
- 17 Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action  
18 observation and acquired motor skills: An FMRI study with expert dancers. *Cerebral*  
19 *Cortex, 15*, 1243–1249.
- 20 Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2006). Seeing  
21 or doing? Influence of visual and motor similarity in action observation. *Current*  
22 *Biology, 16*, 1905-1910.
- 23 Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural  
24 mechanisms of empathy in humans: A relay from neural systems for imitation to limbic

- 1 areas. *Proceedings of the National Academy of Sciences of the United States of America*,  
2 100, 5497-5502.
- 3 Carron, A. V., Widmeyer, W. N., & Brawley, L. R. (1985). The development of an  
4 instrument to assess cohesion in sports teams: The Group Environment Questionnaire.  
5 *Journal of Sport Psychology*, 7, 244-266.
- 6 Chang, V. T., Hwang, S. S., & Feuerman, M. (2000). Validation of the Edmonton symptom  
7 assessment scale. *Cancer*, 88, 2164-2171.
- 8 Chow, G. M., & Feltz, D. L. (2008). Exploring new directions in collective efficacy and  
9 sport. In M. Beauchamp & M. Eys (Eds.), *Group dynamics in exercise and sport*  
10 *psychology: Contemporary themes* (pp. 221-248). New York, NY: Routledge Publishers
- 11 Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- 12 Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York:  
13 Chapman & Hall.
- 14 Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines  
15 for choosing between multi-item and single-item scales for construct measurement: A  
16 predictive validity perspective. *Journal of the Academy of Marketing Science*, 40, 434-  
17 449.
- 18 Dimoka, A. (2011). Brain mapping of psychological processes with psychometric scales: An  
19 fMRI method for social neuroscience. *NeuroImage*, 54, S263-S271.
- 20 Edwards, P., Roberts, I., Sandercock, P., & Frost, C. (2004). Follow-up by mail in clinical  
21 trials: Does questionnaire length matter? *Controlled Clinical Trials*, 25, 1, 31-52.
- 22 Feltz, D. L., & Lirgg, C. D. (1998). Perceived team and player efficacy in hockey. *Journal of*  
23 *Applied Sport Psychology*, 83, 557-564.

- 1 Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus  
2 multiple-item measurement scales: An empirical comparison. *Educational and*  
3 *Psychological Measurement, 58, 6*, 898-915.
- 4 Gibson, C. B. (1999). Do they do what they believe they can? Group efficacy and group  
5 effectiveness across tasks and cultures. *Academy of Management Journal, 42*, 138–52.
- 6 Greenlees, I. A., Graydon, J. K., & Maynard, I. W. (1999). The impact of collective efficacy  
7 beliefs on effort and persistence in a group task. *Journal of Sports Sciences, 17, 2*, 151-  
8 158.
- 9 Greenlees, I. A., Graydon, J. K., & Maynard, I. W. (2000). The impact of individual efficacy  
10 beliefs on group goal selection and group goal commitment. *Journal of Sports Sciences,*  
11 *18*, 451-459.
- 12 Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of  
13 team-efficacy, potency, and performance: Interdependence and level of analysis as  
14 moderators of observed relationships. *Journal of Applied Psychology, 87*, 819-832.
- 15 Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist?  
16 *The Quarterly Journal of Experimental Psychology, 58*, 193-233.
- 17 Heuze, J., Sarrazin, P., Masiero, M., Raimbault, N., & Thomas, J. (2006). The relationships  
18 of perceived motivational climate to cohesion and collective efficacy in elite female  
19 teams. *Journal of Applied Sport Psychology, 18*, 201-218.
- 20 Hoepfner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility  
21 of a single-item versus multiple-item measure of self-efficacy in predicting relapse  
22 among young adults. *Journal of Substance Abuse Treatment, 41, 3*, 305-312.
- 23 Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales:  
24 Methodological artifact or response style? *Structural Equational Modelling: A*  
25 *Multidisciplinary Journal, 10*, 435-455.



- 1 Jordan, J. S., & Turner, B. A. (2008). The feasibility of single-item measures for  
2 organizational justice. *Measurement in Physical Education and Exercise Science, 12*,  
3 237-257.
- 4 Kozub, S. A., & McDonnell, J. F. (2000). Exploring the relationship between cohesion and  
5 collective efficacy in rugby teams. *Journal of Sport Behavior, 23*, 120-129.
- 6 Kwon, H., & Trail, G. (2005). The feasibility of single-item measures in sport loyalty  
7 research. *Sport Management Review, 8*, 69-88.
- 8 Li, F., & Harmer, P. (1996). Confirmatory factor analysis of the Group Environment  
9 Questionnaire with an intercollegiate sample. *Journal of Sport and Exercise Psychology,*  
10 *18*, 49-63.
- 11 Lindsley, D. H., Brass, D. J., & Thomas, J. B. (1995). Efficacy performance spirals: A  
12 multilevel perspective. *Academy of Management Review, 20*, 645-678.
- 13 Loo, R. (2002). A caveat on using single-item versus multiple-item scales. *Journal of*  
14 *Managerial Psychology, 17*, 68-75.
- 15 Marsh-Richard, D. M., Hatzis, E. S., Mathias, C. W., Venditti, N., & Dougherty, D. M.  
16 (2009). Adaptive visual analog scales (AVAS): A modifiable software program for the  
17 creation, administration, and scoring of visual analog scales. *Behavioral Research*  
18 *Methods, 41*, 99-106.
- 19 Miller, J. C., Meier, E., Muehlenkamp, J., & Weatherly, J. N. (2009). Testing the validity of  
20 Dixon & Johnson's (2007) gambling functional assessment. *Behavior Modification, 33*,  
21 156-174.
- 22 Myers, N. D., Feltz, D. L., & Short, S. E. (2004). Collective efficacy and team performance:  
23 A longitudinal study of collegiate football teams. *Group Dynamics: Theory, Research,*  
24 *and Practice, 8*, 126-138.

- 1 Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal*  
2 *of Occupational and Organizational Psychology*, 75, 77-86.
- 3 Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends*  
4 *in Cognitive Sciences*, 10, 59-63.
- 5 Rameson, L. T., Morelli, S. A., & Lieberman, M. D. (2011). The neural correlates of  
6 empathy: Experience, automaticity, and prosocial behavior. *Journal of Cognitive*  
7 *Neuroscience*, 24, 235-245.
- 8 Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem:  
9 Construct validation of a single item measure and the Rosenberg Self-Esteem Scale.  
10 *Personality and Social Psychology Bulletin*, 27, 151-161.
- 11 Roe, D. (2008). Question stem. In P. Lavrakas (Ed.), *Encyclopedia of survey research*  
12 *methods* (pp. 666-667). Thousand Oaks, CA: SAGE Publications, Inc. doi:  
13 <http://dx.doi.org/10.4135/9781412963947.n429>
- 14 Shearer, D. A., Holmes, P., & Mellalieu, S. D. (2009). Collective efficacy in sport: The future  
15 from a social neuroscience perspective. *International Review of Sport and Exercise*  
16 *Psychology*, 2, 38-53.
- 17 Short, S. E., Apostol, K., Harris, C., Poltavski, D., Young, J., Zostautas, N., et al. (2002).  
18 Assessing collective efficacy: A comparison of two approaches. *Journal of Sport and*  
19 *Exercise Psychology*, 24, S115.
- 20 Short, S. E., & Ross-Stewart, L. (2009). A review of self-efficacy based interventions. In S.  
21 D. Mellalieu & S. Hanton (Eds.), *Applied sport psychology advances: A review* (pp. 221-  
22 281). London, UK: Routledge.
- 23 Short, S. E., Sullivan, P., & Feltz, D. L. (2005). Development and preliminary validation of  
24 the Collective Efficacy Questionnaire for Sports. *Measurement in Physical Education*  
25 *and Exercise Science*, 9, 181-202.

- 1 Spink, K. S. (1990). Group cohesion and collective efficacy of volleyball teams. *Journal of*  
2 *Sport & Exercise Psychology, 12, 3, 301-311.*
- 3 Stajkovic, A. D., Lee, D. S., & Nyberg, A. J. (2009). Collective efficacy, group potency, and  
4 group performance: Meta-analyses of their relationships, and test of a mediation model.  
5 *Journal of Applied Psychology, 94, 814-828.*
- 6 Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). *Using multivariate statistics (4<sup>th</sup>*  
7 *ed.)*. Boston, MA: Allyn and Bacon.
- 8 Thom, N. J., Johnson, D. C., Flagan, T., Simmons, A. N., Kotturi, S. A., Van Orden, K. F.,...  
9 Paulus, M. P. (2012). Detecting emotion in others: Increased insula and decreased medial  
10 prefrontal cortex activation during emotion processing in elite adventure racers. *Social*  
11 *Cognitive and Affective Neuroscience, 9, 225-231.*
- 12 Williams, L. J., Cote, J. A., & Buckley, M. R. (1989). Lack of method variance in self-  
13 reported affect and perceptions at work: Reality or artifact? *Journal of Applied*  
14 *Psychology, 74, 3, 462-468.*
- 15 Williams, V. S., Morlock, R. J., & Feltner, D. (2010). Psychometric evaluation of a visual  
16 analog scale for the assessment of anxiety. *Health and Quality of Life Outcomes, 8, 1-8.*
- 17  
18  
19  
20  
21  
22  
23  
24  
25

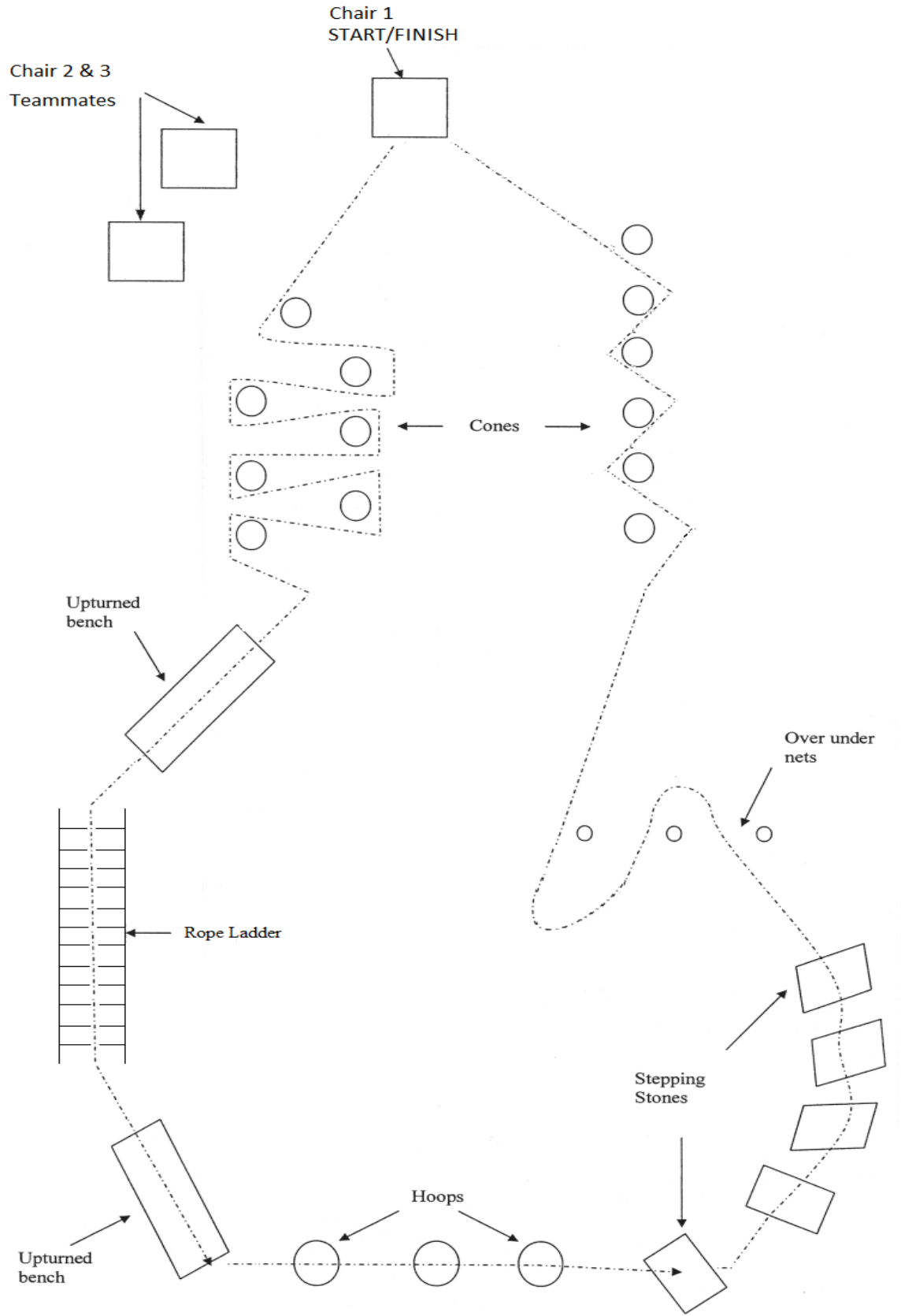
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

### Figure Captions

*Figure 1.* Overhead layout view of obstacle course used for team-based task.

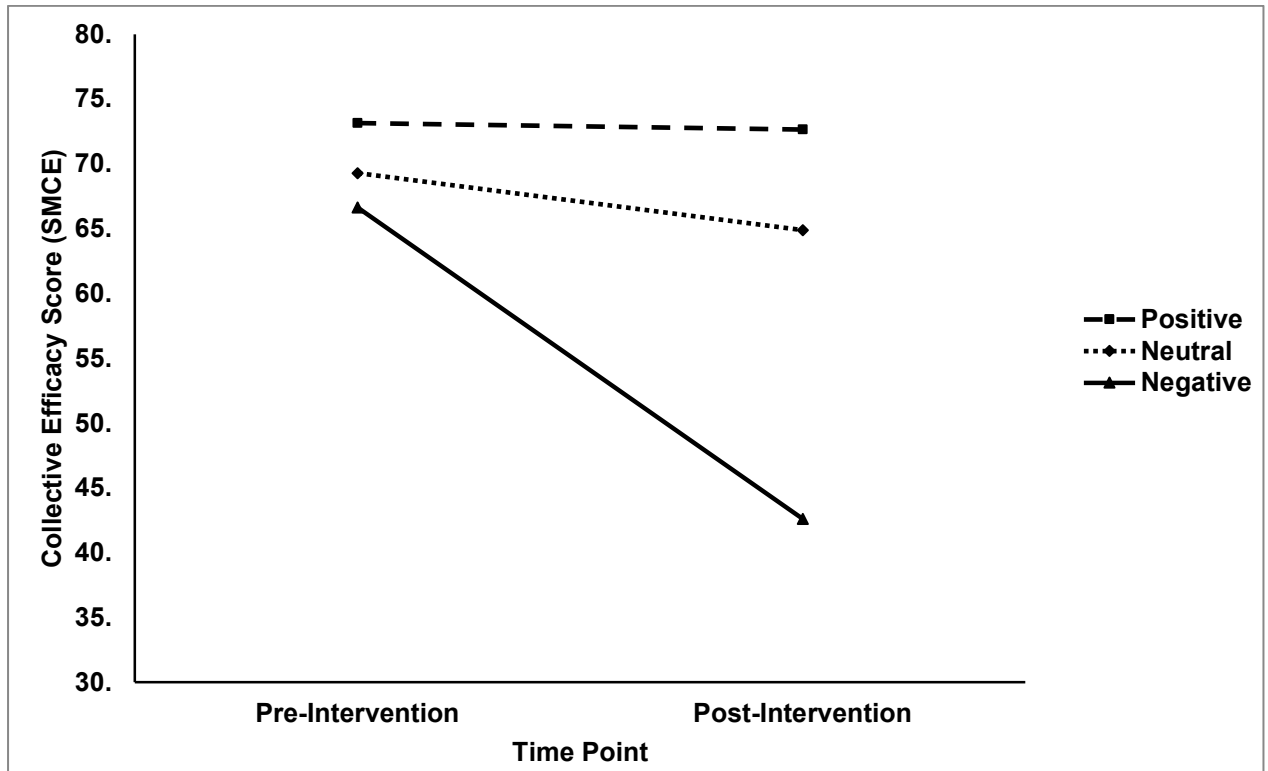
*Figure 2.* Intervention effect upon collective efficacy beliefs for positive, negative and neutral conditions.

*Figure 3.* Intervention effect upon collective efficacy beliefs for basketball and non-basketball conditions.

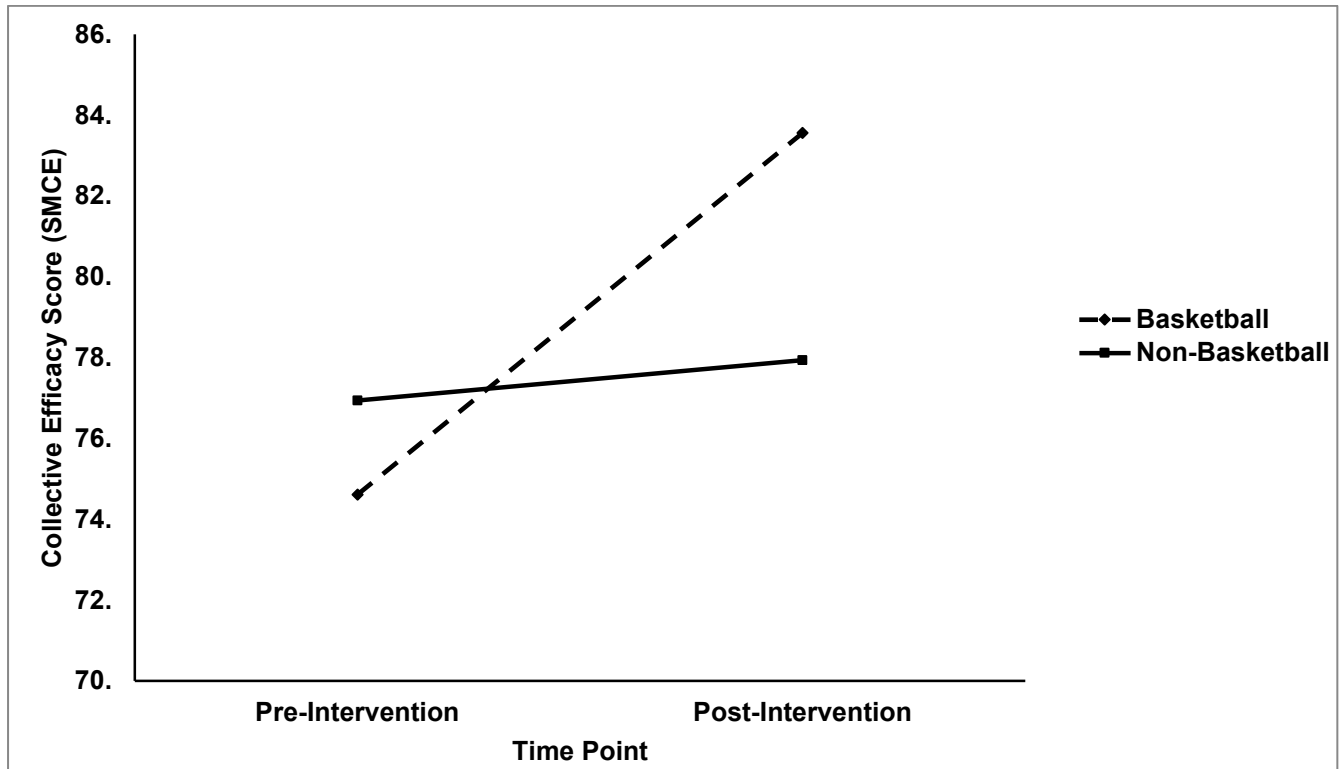


1

2



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13



- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15

1 Table 1  
 2 *Pre- and Post-intervention Collective Efficacy Mean and Standard Deviation for Positive,*  
 3 *Negative, and Control Conditions*

Condition	Pre-intervention		Post-intervention	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
CEQS				
Positive	6.51	.81	7.06	.71
Neutral	6.49	.80	6.79	.92
Negative	6.40	.78	5.44	1.12
Single-item measure				
Positive	73.15	7.85	72.65	9.80
Neutral	69.29	11.75	64.89	13.61
Negative	66.64	14.05	42.60	17.59

4

5

6

7

8

9

10

11

12



1 Table 2

2 *Pre- and Post-intervention Collective Efficacy Mean and Standard Deviation for Basketball*3 *and Non-Basketball Groups*

4

Condition	Pre-intervention		Post-intervention	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
CEQS				
Basketball	6.16	1.08	7.03	1.01
Non-Basketball	6.52	.82	6.81	.87
Single-item measure				
Basketball	74.61	9.91	83.56	6.26
Non-Basketball	76.94	8.34	77.94	8.83

5

6

7

8

9

10

11

12

13

14