Running head: THE COACHING OF PSYCHOLOGICAL SKILLS SCALES

Coaches Coaching Psychological Skills – Why Not? A Framework and Questionnaire Development

Date submitted 06.12.2018

Abstract

The present paper is part of a program of research investigating coaches delivering psychological skills (PS). Here three studies feature an original conceptualization of coaching PS and the development and validation of two questionnaires capturing the coaching of PS. We conducted a qualitative investigation to establish a conceptual framework which included the fundamental coaching of PS behaviors (CPS-F) and the needs supportive coaching of PS (CPS-NS). We then tested the factor structure of two subsequently developed questionnaires via a Bayesian Structural Equation Modelling (BSEM) approach to Confirmatory Factor Analysis across two samples and ran tests of invariance, concurrent, discriminant and predictive validity. The CPS-F questionnaire showed an excellent fit for a three-factor model, whereas the CPS-NS demonstrated an excellent single factor fit. Significant relationships with theoretically related constructs suggested concurrent, discriminant and predictive validity. The findings are expected to significantly further research into our understanding of coaches coaching PS.

Keywords: Coaching, Psychological Skills, Questionnaire, Validation, Bayesian

Prologue

The present paper is part of a program of research arising from the interests of Sport Wales (a UK National Sport Institute) in coaches delivering psychological skills (PS) to their athletes, with the overarching aim of gaining insights into the coaching of PS and developing an effective intervention to upskill coaches in PS. Given the lack of rigorous research testing in this area, the research program was developed and conducted in three phases of research studies following the Medical Research Council guidelines for complex interventions (Craig et al., 2008). The first phase involved piloting the feasibility of a coaching PS intervention based on behavior change theory (i.e., Self-Determination Theory; Deci & Ryan, 2000). From the pilot investigation, it was clear that the intervention had promise, but several adjustments were needed to make the research process and intervention more effective. In particular, we found that the coaching of PS involved a broad set of coaching behaviors that had not been previously documented, and were not adequately captured by measures used in the pilot study. As such, the second phase of the research program involved developing a coaching PS framework and then validating two coaching of PS questionnaires. The third phase of the program was a quasi-experimental controlled trial to evaluate the effectiveness of the adjusted intervention informed by the pilot intervention and evaluated using the validated questionnaires. The pilot intervention and a quasi-experimental intervention trial (Phase 1 and 3) are presented together in another manuscript in preparation ([redacted], in prep; see Supplementary file 1 for a detailed summary of this manuscript). The current paper reports on Phase 2, describing the development of the coaching of PS framework along with creating two questionnaires and examining questionnaire validity.

Introduction

Research demonstrates that psychological skills (PS) benefit athlete performance and well-being (e.g., Weinberg & Comar, 1994). In terms of athletes' PS development, research

has mainly focused on the training provided by sport psychology experts (e.g., Thelwell, Greenlees, & Weston, 2006). However, athletes can also develop PS as result of interactions with coaches and peers (Gould, Dieffenbach, & Moffett, 2002), and coach provision of PS training could offer multiple benefits to athletes. Indeed, coaches who have good relationships and regular contact with athletes could be in an ideal position to help athletes incorporate PS consistently into training. In addition, coaches are far greater in number than sport psychology practitioners and coach delivery of PS training would make PS support available to many more athletes.

Despite the potential advantages, the coaching of PS by coaches rarely occurs, and past research suggests that coaches report a lack of confidence and knowledge as barriers in delivering PS (Callow, Roberts, Bringer, & Langan, 2010; Paquette & Sullivan, 2012). A small number of coach PS interventions exist (Callow et al., 2010; Edwards, Law, & Latimer-Cheung, 2012; Hall, Jedlic, Munroe-Chandler, & Hall, 2007; Hall & Rodgers, 1989; Harwood, 2008), with these typically being workshop based and evaluated via coach self-report. Such interventions have produced some positive outcomes (e.g., positive attitudes towards PS), but coaching behavior has not been rigorously evaluated and often remained unchanged (Edwards et al., 2012; Harwood, 2008). Importantly, there is a paucity of understanding regarding the nature of coaching PS and what it should involve. To date, there has been no systematic examination of coaching PS and therefore there is no evidence-based framework via which to support coaches to engage in PS training. Harwood has been one of the few researchers to publish any behavioral guidelines regarding the coaching of PS (Hardwood, 2008; Harwood, Barker & Anderson, 2015). Whilst these guidelines are practical and have been applied within interventions, they are limited in terms of being evidence or theory-based. Furthermore, no psychometrically valid measures of coaching PS exist which has hindered progress regarding understanding the possible impact of coaching PS and

improving coaching PS interventions. Indeed, the factorial validity of previous coaching of PS measures (e.g., Gould, Damarjian, & Medbery, 1999; Hall & Rodgers, 1989; Jedlic, Hall, Munroe-Chandler, & Hall, 2007) has largely been untested. With these issues in mind, the current manuscript reports on the creation of a framework of coaching PS, along with the subsequent development of a coaching PS measure.

Conceptualization of PS

Despite extensive investigation into PS researchers rarely define the meaning of PS before measuring it, and a functional definition of PS is lacking. Indeed, multiple PS frameworks (e.g., Durand-Bush, Salmela, & Green-Demers, 2001; Smith, Schutz, Smoll, & Ptacek, 1995; Vealey, 1988) often fail to provide clear distinctions between mental skills (e.g., imagery, goal setting) and other cognitions and/or attributes (e.g., confidence, motivation; cf. Arthur, Fitzwater, Roberts, Hardy, & Arthur, 2017). To advance clarity, we propose the definition of PS should be appropriate to the word 'skill': Either an act or task being performed or an indicator of the standard of performing a task, and that improvement of PS is possible with practice (Tremayne & Newberry, 2005). Although we might contend that confidence and motivation can be improved, it is difficult to conceive carrying out "confidence" or being good at "achievement motivation". Therefore we conclude that concepts such as, confidence, self-esteem, achievement motivation, volition (e.g., Vealey, 1988) are better defined as psychological outcomes that are likely to arise from using PS rather than being defined as PS.

Conceptual ambiguity also pervades in coaching PS measurement tools. For example, Paquette and Sullivan (2012) constructed a scale based on the Mental Skills Questionnaire (MSQ; Bull, Albinson, & Shambrook, 2002) which asked coaches to rate how frequently they implemented seven skills into their coaching sessions (e.g., imagery ability, mental preparation, motivation). Unfortunately, the authors did not comprehensively define PS, and

some subscales (e.g., motivation) are not 'skills'. Additional disparity arises as some scales within the MSQ, measure PS ability (e.g., imagery ability) and others assess PS use (mental preparation). Furthermore, the psychometric properties of the original MSQ (Bull et al., 2002) are yet to be documented, and, as the only example of a PS measurement tool subjected to a rigorous validation attempt, the adapted MSQ of coaching PS revealed poor model fit according to conventional criteria (cf. Hu & Bentler, 1999).

To ensure conceptual clarity in the current research program we align with Hardy and colleagues' (Hardy, Roberts, Thomas, & Murphy, 2010; Thomas, Murphy, & Hardy, 1999) proposal that there are basic cognitive-affective PS (i.e., goal setting, imagery, relaxation and self-talk), and more advanced PS which are indictors of ability (e.g., emotional control, automaticity, attentional control). Performers who practice using basic PS will eventually improve their ability with the more advanced PS, which will ultimately influence performance (see Arthur et al., 2017 for evidence of this effect). To provide a foundation for an appropriate coaching PS measure we focused on the coaching of basic PS defined as cognitive-affective skills (i.e., imagery, goal setting, self-talk and relaxation) which can be learnt, practiced and carried out alongside, or in addition to physical sports performance. We selected the four basic skills of imagery, goal setting, self-talk and relaxation to be the focus of the current investigation, as these are the simplest skills which are most frequently referred to in key texts (e.g., Burton & Raedeke, 2008; Weinberg & Gould, 2015) and qualitative investigations (e.g., Hanton, Mellalieu, & Hall, 2004), and thus perhaps the most relevant for coaches to be delivering to their athletes.

Coaching PS Behavior

Alongside a clear definition of PS, a framework of the specific behaviors involved in the coaching of PS is required. Traditionally effective PS training has been proposed as a structured program delivered systematically in a number of stages (Weinberg & Williams,

2010). However, PS training or delivery are perhaps best seen as coaching. Akin to Lyle's (2002) definition of sport coaching, PS development is a complex and contextually specific process consisting of purposeful, direct and indirect, formal and informal activities designed to improve performance. Therefore, effective PS training could involve coaching activities generally defined as "unlocking a person's potential to maximize their own performance. It is helping them to learn rather than teaching them." (Whitmore, 2009, p.8). Here we use the terminology 'coaching' of PS and as such throughout the paper we endeavor to conceptualize the coaching of PS via inductive analysis and then validate measures of the behaviors involved.

With regards to the measurement of such coaching behavior, multiple conceptualizations and scales covering a broad range of coaching behaviors have been developed and validated (e.g., Callow, Smith, Hardy, Arthur, & Hardy, 2009; Chelladurai & Saleh, 1980; Williams et al., 2003). However, to the best of our knowledge, the Coaching Behavior Scale for Sport (CBS-S; Côté, Yardley, Hay, Sedgwick, & Baker, 1999) is the only validated coaching behavior questionnaire to include any aspects of coaching of PS. It includes a mental preparation subscale (e.g., *my coach provides advice on how to perform under pressure*), a goal setting subscale (e.g., *my coach helps me to identify strategies to achieve my goals*), as well as a competition strategies subscale (e.g., *my coach keeps me focused during competition*). However as a general coaching questionnaire the CBS-S is not specific to PS and some key components of coaching PS have been omitted (e.g., coaching imagery). Furthermore, the CBS-S does not differentiate between coaching PS behaviors in different subscales (e.g., encouragement, monitoring and feedback). Indeed, a questionnaire which permits the separate analysis of different behaviors would progress understanding regarding the impact of different coaching PS approaches on athletes.

Therefore, the aim of this research was to create and validate a specific coaching of PS framework and measures of athlete-reported coaching behavior that include a range of coaching PS behaviors allowing for differential analysis for the different behaviors. We sought to create and validate a questionnaire that could be distributed to any athletes who receive coaching, aged 13 and above. We undertook three studies using different samples of athletes in order to ensure the conceptualization and measures would have a broad application. Study 1 involved a qualitative analysis of coaches and athlete interviews (individuals with experience of coaching PS) to gain a more in-depth and structured understanding of the nature of coaching of PS. Study 2 involved questionnaire development using the qualitative findings, and then testing of the questionnaires' factor structure via a Bayesian Structural Equation Modelling (BSEM) approach to Confirmatory Factor Analysis (CFA). Finally, in Study 3 we confirmed the factor structure of the questionnaires, and investigated the questionnaires' discriminant, concurrent and predictive validity, alongside questionnaire invariance.

Study 1. Qualitative study of coaching PS

In order to outline a framework of coaching PS behavior for questionnaire development we analyzed the interview transcripts of coaches and athletes who had experienced coaching of PS through a PS coaching intervention ([redacted] et al., in prep see supplementary file 1 for details).

Method

Participants. Four elite coaches (two females, two males, M_{years} experience coaching = 15.25, *SD* 6.13, two UKCC level 4 qualified, two level 3 qualified) and five elite athletes (two males, three females, M_{age} = 18.0 years, *SD* 1.83, M_{years} experience of the sport = 6.50, *SD* 1.29, two national level, three international level) agreed to be interviewed.

193 **Interview procedure**. Semi-structured interview guides were used, and probes were
194 established *a priori* in order to deepen interviewees' responses to questions if required
195 (Patton, 2002). We piloted the interview guides and made several minor adjustments prior to
196 interviewing study participants. An experienced interviewer conducted the interviews, this
197 interviewer had not been involved in the previous intervention with the participants.
198 At the end of each coach interview the interviewer asked coaches to identify an
199 athlete who had received their coaching of PS. All the interviews were conducted face to
200 face, the coach interviews lasted an average of 90.60 mins (*SD* = 20.40) and the athlete
201 interviews lasted an average of 54.41 mins (*SD* = 8.28). Interviews were recorded,
202 transcribed verbatim and proof-read by the first author. The first author emailed the
203 participants copies of their transcripts and offered them the opportunity to amend their
204 transcripts. Three coaches replied to the email and provided no amendments.
205 **Data analysis.** For the purposes of the current paper as an in-depth analysis of the
206 nature of coaching PS, we analyzed the interview transcripts via hierarchical content analysis
207 (Sparkes & Smith, 2014) using NVivo software. In this analysis we developed themes and
208 categorizations inductively from the data rather than using any pre-determined categories. We
209 identified all the data describing the coaching of PS as meaningful units of analysis and
210 coded these into nodes (*n* =154). We grouped similar nodes together to establish raw themes
211 with internal homogeneity (where all nodes in one theme share meaningful characteristics)
212 and external heterogeneity (the differences between nodes in different themes are clear;
213 Patton, 2002) and then grouped the raw themes into higher order themes to examine their
214 representativeness.
215 In order to increase the creditability and dependability of results (see Biddle,
216 Markland, Gilbourne, Chatzisarantis, & Sparkes, 2001) the second author, with expertise in
217 coach interventions and PS, acted as a "devil's advocate". The additional researcher critically

questioned the analysis (Marshall & Rossman, 1995) by challenging the inclusion of nodes and themes and actively searching for contradictions in the hierarchical model of coaching PS. The first and second author met on three occasions and discussed each raw theme in turn, regularly returning to initial nodes and interview transcripts. During the meetings, we worked collaboratively to resolve issues and refine the model to describe the nature of coaching PS. Note while these interviews are also a feature of [redacted] et al., in prep, the research question and analyses presented here are completely different (see Supplementary file 1 for further details).

Results

Following the content analysis, we identified 20 first level clusters of raw themes. We grouped these into six dimensions under two categories, the *Fundamental coaching of PS* and the *Needs supportive coaching of PS* (see Figure 1 for framework and quotations).

Fundamental Coaching of PS.

Observation. The coaching of PS involved coaches observing athletes' use of PS. The coaches talked about watching athletes' use of PS and noticing how effective it was. Coaches said they listened to how negative athletes were and watched for breaks in pre-performance routines. Coaches also mentioned testing athletes' use of PS by providing challenges and seeing how well they coped.

Targeted cueing of PS. Targeted cueing of PS involved coaches giving athletes instructions of a psychological nature to focus an athlete's attention on helpful stimuli (i.e., instructing an athlete to imagine the action before they attempt it) without necessarily providing any formal explanations surrounding PS. Targeted cueing involved either instructions regarding technique or motivating athletes.

*Instructing using PS cues***.** When giving technical instructions coaches often instructed athletes to focus on a certain cue or key word, for example "explode" when needing to

accelerate quickly at the start of race. Coaches also used imagery-based cues to deliver instructions and describe movements such as “spinning like a vacuum” and “curved like a banana”.

Motivating using PS cues. Coaches also integrated PS cues into their sessions to help motivate their athletes via setting goals for the athletes and using imagery-based descriptions of them achieving their goals. For example, a coach telling an athlete that they could win gold and describing what that would feel like to win.

Instructing to use PS. Coaches directly instructed their athletes to use PS. Instructing to use PS is overtly telling athletes to use PS (e.g., now make sure you do some imagery before your performance) whereas targeted cueing is more covert, meaning as a coach is communicating they will include PS cues such as images (e.g., think about making a shape like a rainbow).

***Reinforcing PS use*.** Coaches and athletes talked about coaches reinforcing athletes’ use of PS, reminding athletes to use PS and regularly repeating instructions about PS.

Needs supportive coaching of PS. In addition to the Fundamental coaching of PS, we identified a more athlete-centered approach to coaching PS. This category involved coaches helping the athletes to understand what PS are and how to use them in a way which would be relevant to them. After establishing the two dimensions of *Providing explanations* and *Seeking athlete involvement*, the parallel between these dimensions and need supportive elements outlined by self-determination theory (SDT) researchers (Markland & Tobin, 2010) became apparent. As such the category was named ‘Needs supportive coaching of PS’.

Providing explanations of PS. Some coaches went beyond giving PS instructions and explained to athletes how to use PS and which helpful outcomes could result from using PS. The coaches also gave advice and answered questions about PS.

Seeking athlete involvement. In order to enhance athlete involvement in PS development, some coaches talked about providing athlete ownership over PS activities and giving choices of PS exercises. Coaches also asked athletes questions and had discussions with the athletes to help them understand their use of PS. Another element of seeking athlete involvement was coaching PS in a way which would be meaningful to the athletes. In particular, a coach talked about finding ways to introduce PS that would be fun and relevant.

Discussion

The results of the hierarchical content analysis suggested six dimensions of coaching PS which we summarized under two categories, the Fundamental coaching of PS (CPS-F) and the Needs supportive coaching of PS (CPS-NS). The CPS-F involved coach directed behaviors within coaching sessions of (a) Observation of PS use, (b) Targeted cueing of PS, (c) Instructing to use PS, and (d) Reinforcing PS use. The CPS-F are general coaching PS activities which indicate the frequency of coaching PS taking place rather than effectiveness when coaching PS. In contrast the CPS-NS involved tailoring the coaching of PS to the individual by (e) Providing explanations and (f) Seeking athlete involvement (refer to Figure 1 for a summary). Therefore, for the purposes of the subsequent study it seemed logical to create two questionnaires, one which captured the fundamentals of coaching PS and another which captured the quality or need supportive nature of coaching PS.

The CPS-F includes instructing, observation and cueing and most models of coaching deem that instructing and providing knowledge of specialized activities or movements is central to the role of a coach (Potrac & Cassidy, 2006), matching our qualitative findings. It has also been readily noted that accurate observation of athletes is integral to effective coaching (Wagstaff, Arthur, & Hardy, 2017). Furthermore, coaching using analogies and cues has garnered attention, particularly in reference to maintaining performance under pressure (Liao & Masters, 2001).

Given the needs supportive nature of the dimensions Providing explanations and Seeking athlete involvement, these dimensions could be placed within the context of SDT research (Deci & Ryan, 2000). SDT is a well-established theory of human motivation which proposes that the satisfaction of an individual's basic needs (autonomy, competence and relatedness) predict the nature of an individual's motivation and autonomous engagement in specific activities. Specifically, SDT research suggests that the provision of need support (in this case provided by coaches) corresponds to increases in an individual's need satisfaction and subsequent motivation and behavior (e.g., Mageau & Vallerand, 2003; Markland & Tobin, 2010). SDT theorists have suggested that need support involves three key elements: structure, autonomy support and interpersonal involvement (Markland & Tobin, 2010). Structure involves helping individuals to develop clear expectations and beliefs that they are able to effectively engage with a task (Jang, Reeve, & Deci, 2010; Markland & Tobin, 2010). Structure support is provided via explanations regarding behavior-outcome contingencies (Silva et al., 2010) and positive feedback regarding progress. Therefore, within this study, the theme of Explanation provision of PS could be described as need supportive and a key component of structure. Autonomy support involves encouraging individuals to engage in tasks for their own reasons and is provided by minimizing pressure, offering choice and acknowledging an individual's perspective (Markland & Tobin, 2010; Silva et al., 2010). Within the current study the dimension of Seeking athlete involvement regarding PS included; giving choices of PS exercises, asking athletes questions about their use of PS, and coaching PS in a way which would be meaningful to the athletes, which could all be described as autonomy supportive behaviors. The content of the CPS-NS is also supported by previous conceptualizations of effective coaching including the individualization of coaching for different athletes (Callow et al., 2009) and autonomy supportive coaching activities (see

Mageau & Vallerand, 2003 for an overview). As such the concept of CPS-NS should assist to enhance the quality and impact that coaching PS can have.

Study 2. Item development and Exploratory Validation of Coaching PS Scale

In this study we created and validated two coaching PS questionnaires to measure the fundamental coaching of PS and the need supportive coaching of PS. The process involved item development, scale refinement and tests of factorial validity with a sample of prominently recreational athletes.

Item Development

Based on the results of the qualitative analysis, we developed the initial questionnaire items for each theme within the fundamental and need supportive coaching of PS (CPS-F 36 items; CPS-NS 19 items). When writing each item, we referred to the direct quotations and used the participants own words whenever possible. We followed widely accepted principles of good practice of questionnaire design whereby we sought to create clearly worded items which asked singular questions and did not contain double negatives (Schwarz, 2007). We also adapted a number of items (n = 5) from Markland & Tobin's (2010) measure of need support. We selected items from Markland and Tobin's questionnaire on the basis that they closely represented the themes found in the qualitative data and these items had previously demonstrated factorial and predictive validity so merited inclusion.

We gave three members of the research team and two additional academic experts in SDT and PS evaluation sheets with a list of all 55 items. We asked each reviewer to conduct an independent review of each item and provide written comments on (a) the clarity of each item, and (b) the relevance of the item to the appropriate theme. We then met as a group and discussed each item in turn whilst considering all written comments relating to each item. We removed items if there were any concerns from reviewers regarding the items' clarity or relevance, and no new items were added. The iterative process of written and verbal feedback

we undertook promoted a depth of analysis of the items and the conceptualization. Indeed, during the review process we established that reinforcing PS use was conceptually distinct from the fundamental behaviors. Specifically, to reinforce PS use with an athlete, PS would have normally been introduced to an athlete at an earlier time and suggests some longevity of coaching PS. As such, reinforcement is relevant to the coaching of PS but it has a different temporal nature to the other behaviors and so it was removed from the questionnaires.

Following this process we were left with two reduced sets of items that we used to create the two measures. The CPS-F questionnaire consisted of 16 randomly ordered items. Participants were asked to rate how frequently the situations occur on a 5-point scale (*0 = never, 1 = rarely, 2= sometimes, 3= often, 4= Always*). The CPS-NS questionnaire consisted of 14 randomly ordered items. Participants were asked to rate their experiences of coaching PS on a 5-point scale (*0 = Not at all true of me, 4 =Very true of me*). The two scales used different anchors on the 0-4rating scales, as '*never to always*' was intended to capture ratings of the frequency of coaching behavior, whereas '*not at all true of me to very true of me*' captured athletes' personal experiences of the coaching of PS when it occurs. All questionnaire items had item stems that were generic and appropriate for all PS 'e.g., my coach instructs me to use ..' with interchangeable subjects for the appropriate PS being measured 'my coach instructs me to use…goal setting' or 'my coach instructs me to use …..imagery.' (see Table 1 for example items from CPS-F and CPS-NS).

Method

Participants. We recruited athletes from Universities and sport clubs who were over the age of 16, received regular coaching (at least one hr. per week) and were actively competing in sport(s). Two hundred and fifty nine athletes agreed to participate (117 males, 142 females, M_{age} = 27.00 years, *SD* 12.54, M_{years} experience of the sport = 9.34, *SD* 7.13). Participants were involved in 34 different sports and responses indicated that, 13.9 % were

competing professionally/internationally, 14.3% nationally, 8.9% regionally, 5.9% in British Universities Leagues, 43.6% recreationally and 13.4% did not report their level of participation.

Data collection procedure. We obtained institutional ethical approval and all participants provided informed consent to participate. There were four versions of the questionnaire each of which referred to a different basic PS. We randomly allocated each athlete to complete one version of the questionnaire (goal setting $n = 68$, imagery $n = 62$, relaxation $n = 59$ and self-talk $n = 70$). We informed the athletes about the purpose of the study, along with information to emphasizing confidentiality, to reduce the potential for social desirability to influence responses on the questionnaire (e.g., we informed athletes that there were no right or wrong answers).

Analyses. There were little missing data (highest 1.9 % missing across CPS-F items and CPS-NS items) and the entire response scale on both measures was used suggesting that the items were sufficiently sensitive to detect differences in coaching received by athletes.

We tested the factor structure of the questionnaires using Bayesian structural equation modelling (BSEM; Muthén & Asparouhov, 2012) which is a novel approach increasingly advocated in the sport and exercise psychology literature (e.g., Myers, Ntoumanis, Gunnell, Gucciardi, & Lee, 2017; Niven & Markland, 2016). The BSEM approach views parameters as variables with a mean and distribution rather than constants, as in a Maximum Likelihood analysis. The BSEM approach allows the researcher to specify more realistic models and simultaneously allow small variances, cross-loadings and correlated residuals within an identified model (see Muthén & Asparouhov, 2012 and also Niven & Markland, 2016 for a detailed overview) which results in more appropriate model fit statistics.

In line with contemporary procedures (e.g., Myers et al., 2017; Niven & Markland, 2016) we first standardized the data and then estimated a series of three BSEM models. The

first model incorporated non-informative priors for the major loadings, exact zero cross-loadings and exact zero residual correlations. The second model incorporated the addition of informative approximate zero cross-loadings. The final model incorporated the addition of both informative approximate zero cross-loadings and residual correlations. We specified the priors with a mean of 0 and a variance of .01. This size of prior corresponds to factor loadings and residuals with a 95% limit of ±.20, therefore representing small cross-loadings and correlated residuals (Muthén & Asparouhov, 2012; Niven & Markland, 2016). We estimated all BSEM models with the Markov Chain Monte Carlo (MCMC) simulation procedure with a Gibbs sampler and a fixed number of 100,000 iterations for two MCMC chains. This procedure allowed for the examination of model convergence.

We assessed model convergence with the potential scale reduction factor (PSR). Model convergence is evident when the PSR value lies between 1.0 and 1.1 for all parameters (Gelman, Carlin, Stern, & Rubin, 2004). In addition, we performed a visual inspection of trace plots for each parameter to check that the parameter values in each MCMC chain mixed well (i.e., converged to a similar target distribution; van de Schoot & Depaoli, 2014). We assessed model fit using the posterior predictive *p* value (PP*p* value). A good-fitting model is indicated when values are around .50 (Muthén & Asparouhov, 2012). In addition, we also examined the symmetric 95% credibility interval for the difference between the observed and replicated χ^2 values. A good fitting model is indicated when the values center on zero (Muthén & Asparouhov, 2012). Once the final models were established we performed a sensitivity analysis to examine if the specification of different prior variances influenced the posterior predictive *p* value and the variability of the estimates (Muthén & Asparouhov, 2012). To do this we reran the final models with variance priors specified at .005, .01 and .015 for the cross-loadings, and then examined parameter estimates to check for any important discrepancies.

Results and Discussion

CPS-F. The 16-item model achieved convergence and all factor loadings were significant. However, the PP*p* indicated an unacceptable fit to the data (See Table 1 for PP*p* and 95% credibility intervals). To improve model fit we considered items for removal based on theoretical relevance and low factor loadings and subsequently removed four items. Such a removal process is common and accepted in measurement development provided that any removals are made based on theory *and* relevant data or evidence, as opposed to simply relying on a data driven approach (e.g., Biddle et al., 2001; Markland, 2007). We removed the Observation item "*My coach watches out for my use of [specific PS] during my sport*" as it was thought of as ambiguous as 'watching out' could mean that a coach deliberately observes PS use, but it could also be interpreted as a coach protecting and looking after an athlete's PS use. This item also had a low factor loading in comparison to the other items. In addition, we removed the Observation item "*my coach tests my use of [specific PS]*" as testing use of a skill is not observing and the Instruction item "*my coach asks me to use [specific PS]*" as this item was thought to be overly similar to another, more specific item "*My coach asks me to think about using [specific PS] when I'm doing my sport*." included in the scale. We also removed the Instruction item "*my coach instructs me to focus [use specific cue] whilst doing my sport*" because, in comparison to the other items in the scale, it was overly different across versions (goal setting, imagery, relaxation, self-talk).

Following this item removal process we analyzed the fit of the 12-item model with and without small variance priors on the cross loadings. The model with non-informative priors failed to converge. The model with informative priors on the cross loadings achieved adequate convergence (with final PSR values below 1.1) yet the fit was still poor (see Table 1). One Instruction item "*My coach tells me to think about [specific cue] when I am performing my sport*" wanted to cross load on targeted cueing beyond its a priori limits. We

deemed the item to be overly close to cueing and subsequently removed it from the model. This process resulted in an 11-item scale with three subscales: observation (n_{items} = 3), targeted cueing (n_{items} = 5), and instruction (n_{items} = 3).

All 11-item BSEM models achieved adequate convergence. The PP*p* for the model with non-informative priors indicated a less than desirable fit to the data. The PP*p* for the model with informative small variance priors on cross-loadings indicated an improved fit (in comparison to the model with no priors), but the resulting fit was still poor. The PP*p* of .53 indicated excellent fit for the final model with informative small variance priors on cross-loadings and residual correlations. In addition, the 95% posterior predictive credibility intervals centered on zero (See Table 1).

All major loadings in the 11-item scale were significant (See Table 1 for standardized factor loadings and 95% credibility intervals for the 11-item scale). PSR values for the final model reached the convergence criterion at 11800 iterations and visual inspection of the trace plots showed support for convergence (i.e., all plots showed a stable convergence across iterations for the two chains). Interfactor correlations (and 95% credibility intervals) were as follows: Targeted Cueing with Observation = .66 [.49, .79], Targeted cueing with Instruction = .68 [.51, .80], Instruction with Observation = .88 [.77, .99]. Further, sensitivity analyses revealed stable factor loadings and cross loadings when specifying larger (.015) and smaller (.005) variance priors. Indeed, 100% of all discrepancies were within ±.05. Composite reliability coefficients (Fornell & Larcker, 1981) for the three subscales were: Observation 0.93, Targeted cueing 0.92, and Instruction 0.93. The constructs of instruction and observation are conceptually distinct (a coach could instruct an athlete to do something without observing them), however the strength of correlation between them led to us re-analyzing the data as a two factor model with targeted cueing as one factor, and Instruction and Observation combined as a single factor. This two factor model also revealed an

excellent fit to the data PP*p* = .54 [-36.17, 33.30], thus from a measurement perspective at least it does not seem to matter with Instruction and Observation are considered separately or as one factor. However, from a conceptual perspective we contend that they are best thought of as two related, yet separate, behaviors.

CPS-NS. The initial 14-item CPS-NS with non-informative priors reached convergence but revealed a poor fit to the data (see Table 1). To improve model fit, we removed three Explanation Provision items ("*my coach suggests ways I could use [specific PS]*", " *my coach explains how to use [specific PS] effectively*", "*my coach provides me with positive feedback about my use of [specific PS]"*) and two Seeking Involvement items *(" my coach asks me questions about my use of [specific PS]*" and "*my coach encourages me to reflect on my use of [specific PS]"*) based on theoretical reasoning. We felt these items failed to sufficiently describe need supportive coaching to its fullest extent as, for example, a coach could ask questions or suggest ways to use a particular PS in a controlling manner. In addition, the item "*my coach provides with me positive feedback*" was conceptually distinct from the other explanation items as it did not refer to explanations about PS use.

Following item removal, we next tested this 9-item model with non-informative priors and then with informative priors on the cross loadings. Both of these models revealed very poor fits but no items wanted to cross load above their accepted limits in the second of these two analyses. We subsequently examined the fit of 9-item model with informative priors on cross loadings and residuals correlations. This analysis resulted in an excellent fit although the residual for one involvement item ("*my coach talks to me about [specific skill] in a way which is relevant to me*") correlated with an explanation provision item beyond its accepted limits. Because this item could conceivably be considered as explanation provision we subsequently removed this item leaving an 8-item model (see data in Table 1).

We then tested the 8-item model with the three BSEM models. All models converged although the models with non-informative priors and with informative priors on the cross loadings revealed poor fits. However, the model with informative priors on cross loadings and residuals revealed an excellent fit, and no items had problematic cross loadings or correlated residuals. All major loadings for items in the 8-item model were significant (see Table 1 for standardized factor loadings). PSR values reached the convergence criterion at 5000 iterations and inspection of trace plots showed support for convergence. The correlation between the two factors (Seeking athlete involvement and Explanation provision) was .96 [.90, .99]. Sensitivity analyses again revealed stable factor loadings and cross loadings at different levels of prior, with 100% of all discrepancies again within ±.05. Composite reliabilities for the two subscales were 0.96 (Explanation Provision) and 0.94 (Seeking Athlete Involvement).

Although the BSEM analyses supported the two-factor structure of the CPS-NS, the correlation between the two factors was substantial. Consequently, we re-analyzed the data as a "true" single factor model. Here, all items loaded onto one factor, to examine, from a measurement perspective, whether the two factors were better replaced by a single need support factor. The true single factor model revealed an excellent fit to the data (PP*p* = .52, 95% CIs [-27.10, 25.24] and had a similar Deviance Information Criterion (3386.29) to the two factor model (3385.60) indicating that both models are equally appropriate. Consequently, while explanation provision and seeking athlete involvement are theoretically distinguishable constructs they do not appear distinguishable at a measurement level.

In summary, after utilizing the BSEM approach and deleting several items based on conceptual and empirical grounds, the final CPS-scales consisted of a three factor 11-item measure of CPS-F (Observation, Targeted cueing, and Instruction) and a single factor 8-item measure of CPS-NS (Explanation Provision and Seeking Athlete Involvement), both with

good model fits. The CPS-F and the CPS-NS are the first psychometrically validated measures of coaching of PS. Furthermore, rather than being a global scale, different behaviors are measured by different subscales. As such, researchers and practitioners are now able to differentiate between the fundamentals of coaching PS and the quality of need supportive nature of coaching PS. Interested readers are directed to the Supplementary file 2 Table S1 detailing the mean and standard deviations for each coaching behavior and PS from the present study.

Study 3. Confirmatory validation of coaching PS Scale

In this study we confirmed the factor structure of the two coaching PS questionnaires (CPS-F and CPS-NS) following the same BSEM approach used in the previous study, but with a different sample of younger, more elite level athletes. Within this study we tested the concurrent, discriminant and predictive validity of the new questionnaires and also examined approximate measurement invariance.

We examined the concurrent validity of the CPS-F and CPS-NS by conducting correlations between the coaching of PS and the coaching of mental preparation using the CBS-S (Côté et al., 1999). A key purpose of PS training or the coaching of PS is assisting athletes with their mental preparation (Weinberg & Williams, 2010). Thus, we hypothesized that all subscales measured by the CPS-F and CPS-NS would be significantly correlated with athletes' ratings of coaching mental preparation on the CBS-S.

To evaluate the discriminant validity of the two PS questionnaires, we explored the questionnaires' ability to discriminate between (a) athlete performance level and, (b) the coaching qualification attained by their coach. We hypothesized that the measures would discriminate between athletes of different performance levels, with higher level athletes reporting more coaching of PS (e.g., Jedlic et al., 2007). Further, we also expected the measures to discriminate between levels of coaching qualification, in that coaches with

higher coaching qualifications would coach more PS than those with lower coaching qualifications (e.g., Hall et al., 2007).

In relation to predictive validity, we expected that the coaching of PS would impact positively on athletes' PS use. However, before increasing PS use, the development of athlete awareness surrounding PS is proposed as a necessary first step towards more effective PS use (Weinberg & Williams, 2010). Therefore, the regular coaching of PS should primarily predict athletes' awareness and knowledge about their PS use before the effective application of PS. Self-awareness is readily mentioned in applied sport psychology literature (e.g., Ravizza, 2010), but has not been empirically measured or investigated in this research field. However, within educational research the concept of awareness as metacognition, which is viewed as an "awareness and management of one's own thought" (Kuhn & Dean, 2004, p.270) has been investigated in some depth. Metacognition is thought to be made up of a multidimensional set of cognitive skills, much like PS in sport (Kuhn & Dean, 2004). Schraw and Dennison (1994) suggested that metacognitive awareness is made up of an individual's knowledge of his/her own cognition and their regulation of cognition. For the purposes of the present research, we were interested in athletes' knowledge of their PS use as a form of awareness and how coaching of PS as measured by the CPS-F and CPS-NS impacts on it. The three types of knowledge as measured by the Mental Awareness Inventory (MAI; Schraw & Dennison, 1994) have been defined as (a) procedural knowledge of cognition, knowledge about how to implement mental strategies (b) declarative knowledge of cognition, knowledge of one's skill and ability to use PS, and (c) and conditional knowledge of cognition, knowledge about when and why to use PS. These three aspects of metacognitive knowledge are thought to be affected by the teaching an individual receives (Schraw & Dennison, 1994). As such, we hypothesized that the coaching of PS (all subscales of the CPS-F and the CPS-NS) would be correlated with a global score from the MAI for sport (MAI-S). We also expected that CPS-

NS would account for significant variance within mental awareness over and above that accounted for by CPS-F. Indeed, providing explanations and involving the athletes in developing PS, was expected to engage the athletes and therefore develop their awareness to a greater extent, than the fundamental coaching PS behaviors.

Finally, we examined approximate measurement invariance in both of the measures (cf. Muthén & Asparouhov, 2013) across the four PS (goal setting, imagery, relaxation and self-talk) by testing the factor structure (configural invariance) and factor loadings (metric invariance).

Method

Participants.

We recruited athletes aged 13 and over from sport teams/clubs and Universities who received regular coaching (more at least one hr. a week) and were actively competing in sport(s). Four hundred and fifty five athletes agreed to participate (257 male, 198 female, M_{age} = 17.69, *SD* 5.22, M_{years} experience of the sport =7.41, *SD* 4.25). Participants were involved in 20 different sports and responses indicated that, 30.1% were competing professionally/internationally, 38.9% nationally, 10.1% regionally, 8.8% in British Universities Leagues, 6.4% recreationally, and 5.7% did not report their level of participation.

Measures.

Coach Behavior Scale for Sport (CBS-S). The CBS-S (Côté et al., 1999) mental preparation subscale included five items which examined coaching behavior to help athletes mentally prepare for their sport (e.g., My coach provides advice on how to perform under pressure). Athletes scored all items on 1-7 Likert-type scale (*1 = Never, 2 = Very rarely, 3 = Rarely, 4 = Fairly often, 5 = Often, 6 = Very Often, 7 = Always*). The factor structure of the CBS-S has been explored (Côté et al., 1999) and in the current study, a BSEM of the CBS

590 with correlated residuals revealed an excellent fit (PP*p* = .50, [-17.99, 17.55] and good
591 composite reliability; 0.86.

592 ***Metacognitive Awareness Inventory for Sport (MAI-S).*** We adapted The MAI
593 (Schraw & Dennison, 1994) subscale measuring an individual's knowledge of his or her own
594 cognition to apply to a sports context (E.g., *I am aware of what strategies I use when I study*
595 was adapted to *I am aware of what mental strategies I use when I play sport*). Each item was
596 rated against a 100mm, bipolar scale, the right end labelled *true* and the left end *false*, and
597 participants recorded their responses by drawing a line across the scale. The length of the
598 length was measured in mm and was then reverse scored. Previous factorial analyses have
599 been conducted on both the MAI (Schraw & Dennison, 1994) and Junior MAI (Sperling,
600 Howard, Miller, & Murphy, 2002) suggesting variable model fit. BSEM analyses revealed
601 the MAI-S had a 3-factor, 12-item scale to have an excellent fit (PP*p* = .51, [-38.51, 37.97])
602 which revealed acceptable composite reliability (procedural knowledge α = .82, declarative
603 knowledge α = .79, conditional knowledge α = .75). A copy of the adapted MAI-S
604 questionnaire can be found in the Supplementary file 5.

605 **Data collection procedure.** Following institutional ethical approval, all participants
606 provided informed consent. For any athletes under 16 the adult in care of the young person
607 provided consent. Two hundred and seventy-six athletes were randomly allocated to one
608 version of the CPS-F and CPS-NS questionnaire to complete. The data from a further 179
609 athletes from ([redacted] et al in prep) were used. Thus, in total the number of questionnaires
610 completed were as follows: goal setting n =129, imagery n = 105, relaxation n = 106 and self-
611 talk n = 113. We informed the athletes about the purpose of the study and gave anti-social
612 desirability instructions to emphasize confidentiality. With permission from national
613 governing bodies and coaches, we collected the data at sport training and competition venues.
614 Whilst all athletes completed the CPS-F and CPS-NS, sub-samples also completed the CBS-S

(*n* = 271, M_{age} = 18.4 *SD* 3.8, *n* = 150 male, *n* = 121 female) and the MAI-S (*n* = 371, M_{age} = 17.34 *SD* 5.3, n = 215 male, n = 156 female).

Analyses and Results

Preliminary analysis revealed very few missing data (highest 3.9% missing across all CPS-F and CPS-NS items) and the entire response scale on both measures was used suggesting that the items were sufficiently sensitive.

Factor structure of CPS-F and CPS-NS. We used the same 3-stage BSEM approach from Study 2 to examine the model fits of the 11-item CPS-F and the 8-item CPS-NS. For both measures, the models with non-informative priors and informative priors on cross loadings only revealed less than acceptable fits. However, the fits of the models including informative priors on the cross loadings and correlated residuals were excellent. The final CPS-F model converged after 31800 iterations and the final CPS-NS model after 9000 iterations. All major factor loadings were significant and similar to those in Study 2 (CPS-F factor loadings ranged 0.93-0.73; CPS-NS factor loadings ranged 0.88-0.81), and neither of the final models had cross loadings or correlated residuals that wanted to load beyond accepted limits. Sensitivity analyses also supported the stability of all parameter estimates for each measure. Correlations between the CPS-F factors were: Targeted Cueing with Observation = .66 [.49, .79], Targeted cueing with Instruction = .68 [.53, .82], Instruction with Observation = .85 [.73, .99], with the correlation between the two factors of the CPS-NS being .96 [.89, .99].

The findings between the two CPS-NS factors (i.e., high correlation) mirrored the results from Study 2. Therefore, we again ran a true single factor model and compared this to the two factor model. The fit of the single factor model was again excellent (PP*p* = .51 [-27.04, 25.38]) with the Deviance Information Criterion (6784.74) being almost identical to the two factor model (6784.49). These findings confirm Study 2 in terms of the two CPS-NS

factors being difficult to distinguish at a measurement level despite being conceptually distinct. Full BSEM data from Study 3 is available upon request from the first author. Interested readers are directed to Supplementary file 2 Table S1 detailing the mean and standard deviations for each coaching behavior and PS from the present study. Final copies of the questionnaires can be found in Supplementary file 4.

Concurrent Validity. We examined the concurrent validity of the CPS-F and CPS-NS via bivariate correlations between CBS-S scores and scores on the CPS-F and CPS-NS subscales. All scales were significantly correlated (see Table 2).

Discriminant Validity.

Performance Level. Discriminant function analysis (DFA) indicated that athletes' reports on CPS-F discriminated between athletes' performance levels, Wilks' $\Lambda = .94$, χ^2 (6, n = 428) = 26.77, $p < 0.001$. The standardized structure coefficients for the first discriminant function revealed that coach instruction of PS ($r = .87$) made the greatest contribution to the discriminant function, followed by targeted cueing ($r = .82$) and coach observation of PS ($r = .56$). Examination of the discriminant function at the group centroids revealed that elite level athletes (.25) reported most fundamental coaching of PS behaviors which discriminated them from lower performing athletes, both competitive athletes (-.03) and recreational athletes (-.56). Athletes' reports on CPS-NS also discriminated between athlete performance levels, Wilks' $\Lambda = .96$, χ^2 (4, n = 418) = 16.96, $p < 0.001$. The standardized structure coefficients suggested that providing explanation ($r = .99$) made the greatest contribution to the discriminant function, followed by seeking athlete involvement ($r = .98$). Examination of the discriminant function at the group centroids revealed that elite level athletes (.24) reported the most coaching of PS which discriminated them from lower performing athletes, both competitive athletes (-.04) and recreational athletes (-.43).

Coaching qualification. For both the CPS-F and CPS-NS, the DFAs were non-significant, indicating that neither measure was able to discriminate between coach level UKCC of equivalent (Group1 = qualification level 1 & 2; Group2 = qualification level 3 & 4). CPS-F Wilks' $\Lambda = .98$, χ^2 (3, $n = 280$) = 6.65, $p = 0.08$; CPS-NS, Wilks' $\Lambda = 1.00$, χ^2 (2, $n = 277$) = 1.04, $p = 0.59$.

Predictive Validity. All factors of the CPS-F and CPS-NS were significantly correlated with athlete awareness of mental strategies on the MAI-S (see Table 2). To determine the extent to which CPS-NS predicts variance in awareness of PS beyond that explained by CPS-F, we conducted a hierarchical regression analysis with the CPS-F variables entered in the first step and the CPS-NS subscales entered at Step 2. The CPS-F variables accounted significant variance in the MAI-S, $R^2 = .09$, F (3, 360) = 12.00, $p < .001$. Moreover, the CPS-NS variables accounted for significant variance over and above that accounted for by the CPS-F, $R^2 = .04$, F (2, 358) = 10.97 $p < .001$. The beta coefficients revealed the unique variance in block two was attributed to Seeking athlete involvement $\beta = .27$ $p = .04$, whereas the beta coefficient for Providing explanations was not significant $\beta = .15$ $p = .26$.

Invariance testing. We estimated all BSEM with MCMC simulation procedure with a Gibbs sampler and a fixed number of 100,000 iterations for two MCMC chains (Gelman et al., 2013). For the correlated residuals we specified an inverse-Wishart prior distribution *IW* (0, degrees-of-freedom parameter d) with $d = p + 20$. We varied three different levels of approximation by specifying zero mean small variance priors of .05, .01 and .005 on the factor loadings (metric invariance). We used the fit indices previously outlined and used the deviance information criteria (DIC) to compare BSEM and any parameters that differed significantly from the priors between PS.

***CPS-F*.** The model for approximate measurement invariance across PS within the CPS-F failed to converge. The non-convergence is most likely a result of an overly complex model (three CPS-F factors across four PS) for the sample size. To overcome this problem, we instead collapsed the observation and instruction factors based on empirical compatibility and ran a two-factor approximate measurement invariance analysis. We maintain that observation and instruction are conceptually distinct constructs but have combined them here simply to reduce model complexity in order to test invariance. Fit statistics are displayed in Supplementary file 3, Table S2. The test for configural invariance indicated excellent fit. The test for approximate metric invariance (factor loadings) resulted in good fit at all prior distributions (.01, .005 and .005) and the DIC statistic showed support for a more parsimonious model at a prior distribution of .005. Further, the Mplus output indicated that there were no invariant parameters for the factor loadings.

***CPS-NS*.** Fit statistics are displayed in Table S2. The test for configural invariance indicated excellent fit. The test for approximate metric invariance (factor loadings) resulted in good fit at all prior distributions (.01, .005 and .005) and the DIC statistic showed support for a more parsimonious model at a prior distribution of .005. Further, the Mplus output indicated that were no invariant parameters for the factor loadings.

Discussion

In Study 3 we confirmed the model fit of the two coaching PS questionnaires, an 11-item CPS-F scale and an 8-item CPS-NS scale using the same BSEM approach as in Study 2 with a different sample. However, as with the first sample, the two CPS-NS factors did not distinguish at a measurement level despite being conceptually distinct. This finding is consistent with other measures of need support in the SDT literature, where different aspects of need support and need satisfaction are routinely collapsed into single scales due to high interfactor correlations (e.g., Markland & Tobin, 2010) but are analyzed as separate

constructs. Indeed, the two needs supportive coaching PS subscales seemed to have different predictive properties based on our other assessments of validity.

We also provided support for the concurrent, discriminant and predictive validity of the CPS scales. All coaching PS subscales correlated with the coaching of mental preparation on the CBS-S (Côté et al., 1999). The CPS-F and CPS-NS discriminated between athletes of different performance levels. Specifically, the elite level athletes reported receiving more coaching of PS, in comparison to competitive or recreational athletes, supporting previous findings (Jedlic et al., 2007). Conversely, there were no differences found between the level of coach qualification and athletes' reports of both fundamental and need supportive coaching of PS. The lack of differences between coaches of different qualification levels and coaching of PS has been found in other research (Hall et al., 2007), this could be due to limited coverage of coaching PS within current qualification training along with the limited effectiveness of formal coaching PS education (Callow et al., 2010).

As an indication of predictive validity, all factors of the CPS-F and the CPS-NS were significantly correlated with athlete awareness of PS, as measured by the MAI-S. This result suggests that coaches observing PS use, providing cues, instructing use of PS and providing needs support regarding PS is related to athletes' knowledge about: how to implement PS, their ability to use PS, and when and why they should use PS. However, the correlations between the variables although significant were relatively small. The strength of correlations may have been due to most coaches in the sample not having been trained in how to coach PS effectively, thus weakening the impact on athletes' awareness of PS. Furthermore, the results of the hierarchical regression indicated that Seeking athlete involvement accounted for unique variance in mental awareness over and above the CPS-F subscales. This result suggests that need supportive coaching behaviors are more influential on athletes' mental awareness, justifying the use of separate questionnaires. This could be explained as seeking

738 athlete involvement (e.g., *My coach encourages me to take my own initiative*) requires more
739 cognitive processing from athletes rather than simply receiving instructions or a coach
740 observing you.

741 The tests of invariance broadly revealed support for approximate metric invariance for
742 the CPS-F and the CPS-NS. However, it is important to note that we were only able to test a
743 two-factor model for the CPS-F due to issues with model convergence. This issue
744 notwithstanding, the important result from the invariance tests was that the factor-loadings
745 were equivalent across the PS in both measures suggesting that the items are good indicators
746 of the underlying latent variable regardless of the skill being assessed.

747 **General Discussion**

748 There has been a lack of rigorous investigation into sport coaches delivering PS. In
749 the current set of studies we offered a definition of basic PS as cognitive-affective skills (i.e.,
750 imagery, goal setting, self-talk and relaxation) which can be learnt, practiced and carried out
751 alongside, or in addition to, physical sports performance. We then developed an original
752 conceptual framework of what coaching PS might involve, along with a novel and
753 psychometrically sound instruments to capture the coaching of PS.

754 Interviews with athletes and coaches who had experience of PS provided the basis for
755 the framework of coaching PS and the two questionnaires developed in this study. The
756 behaviors within the current conceptualization are specific to coaches endeavoring to enhance
757 the basic PS of their athlete, with more in-depth psychological assessment and support under
758 the remit of those with formal training and qualifications (e.g., Health and Care Professional
759 Council registered Sport Psychologists in the UK).

760 As the first concerted attempt to operationalize the coaching of PS that satisfies
761 multiple forms of validity. The CPS-F (11 items) and CPS-NS (8 items) demonstrated a
762 consistently good factor structure across two different samples suggesting that the

questionnaires can be distributed to any athletes receiving coaching and participating in any sport at any level, who are aged 13 and above. We found support for measurement invariance and provided evidence of discriminant validity as elite athletes reported significantly more coaching of PS than lower level athletes. Supportive evidence concerning the concurrent validity and predictive validity was also found. Such results suggest the CPS-F and CPS-NS are meaningful to athletes and that the scores derived are valid and reliable indicators of coaching PS. Furthermore, we adapted and explored the factorial validity of the MAI-S which will be a helpful measure of athlete awareness for sport psychology practitioners and researchers. An interesting point to note in the CPS-F specifically is that instruction and observation correlated substantially across both samples. Indeed, in Study 2 a CFA of a 2 factor solution also revealed a fit that was comparable in quality to the original 3 factor solution. Such findings might cast doubt as to the benefits of separating instruction and observation as coaching behaviors. However, we believe that keeping these as separate behaviors is important for two reasons. First, the DFA's in Study 3 showed that these two behaviors contributed differently to discrimination between groups. Such a result supports the view that these two constructs are best considered as separate, as one would lose important information such as this if these factors were combined. Second, from an applied perspective, separating out these constructs also appears important. If one is conducting an intervention around observation, having a scale that measures this construct (as opposed to a scale that measures a combination of observation and instruction) is likely to yield much more useful information about the benefits of an intervention.

An important strength of the CPS-F and CPS-NS, is that the two questionnaires when used together give an indication of the quantity of coaching behavior and quality of coaching behavior, two aspects which are rarely considered in tandem when capturing coaching behavior. Furthermore, the five different behavioral subscales provide a differentiated

understanding regarding the most effective approach to coaching PS and the effectiveness of coaching PS interventions.

However, due to the difficulty in finding participants with experience of coaching PS, there could have been weaknesses in the conceptualization phase we undertook. Indeed, all the athlete interviewees had been part of a specific coaching PS intervention and were aged between 16 and 21, as such their views regarding the nature of coaching PS could be biased or overly narrow. However, this issue could be somewhat negated given that the coaches we interviewed had experience of coaching PS to a broad range of athlete ability and age groups. Indeed, the findings do mirror the extant coaching literature (Liao & Masters 2001; Potrac & Cassidy, 2006; Wagstaff et al., 2017) that readily identifies instruction, observation and the use of cues, along with the provision of feedback and individualized approaches as vital components of the coaching process. However, as the field progresses and coaching PS becomes more commonplace, there are likely to be other behaviors which will emerge and warrant inclusion in the CPS-F and CPS-NS.

It could be argued that, given our definition of PS, we have limited the boundaries of PS coaching at the expense of conceptual breadth. However, in an attempt to avoid previous conceptual ambiguity and provide clarity, we purposefully offered a tight definition of coaching of PS, with the measurement tool designed in a way that it can be used flexibly to measure other more advanced PS (e.g., attentional control) or multidimensional aspects of PS by changing the stem descriptors (e.g., “My coach gives me good advice about goal setting” could become “my coach gives me good advice about process goals”). Indeed, the questionnaire demonstrated good factorial validity across two samples, which included four different versions of the questionnaires (the coaching of goal setting, imagery, relaxation and self-talk). Thus our definition of coaching PS, coupled with nature of the measurement tools, provides a foundation for future work to develop a more fine-grained understanding of

coaching PS and associated mechanisms. Similarly, the current questionnaires measure the coaching of PS at a general level only, however the coaching of PS occurs in different contexts (e.g., training, pre-match, in-match, post-match), and this would be a worthy consideration in future research.

It is important to note that despite the encouraging results, two factor analytic studies only do not offer complete validation and, indeed, the factor structure should be replicated across different samples with different characteristics. Further replication would be helpful in light of us being unable to test the measurement invariance of the full 3-factor CPS-F. Furthermore, the predictive validity of the questionnaire would be best assessed using longitudinal designs. Nevertheless, the evidence presented in this paper suggests that the CPS-F and CPS-NS will serve as useful tools in future research and the framework of coaching PS could be an essential reference for practitioners developing PS training with coaches. This work significantly progresses understanding of coaching PS and will further the quality of research investigations into coaching PS.

Acknowledgements

We would like to thank Sport Wales and the School of Sport, Health and Exercise Sciences for funding this research, Dr Joy Bringer for her support during the project, and Fiona Meikle and Michelle Smith for their assistance with data collection.

References

Arthur, R., Fitzwater, J., Roberts, R., Hardy, J., & Arthur, C. (2017). Psychological skills and "the Paras": The indirect effects of psychological skills on endurance. *Journal of Applied Sport Psychology 29,* 449-465. doi:10.1080/10413200.2017.1306728

[redacted]

Biddle, S.J.H., Markland, D., Gilbourne, D., Chatzisarantis, N.L.D., & Sparkes, A.C. (2001). Research methods in sport and exercise psychology: quantitative and qualitative issues. *Journal of Sports Sciences, 19*, 777-809. doi:10.1080/026404101317015438

Bull, S.J., Albinson, J.G., & Shambrook, C.J. (2002). *The mental game plan: Getting psyched for sport*. Brighton, UK: Sport Dynamics.

Burton, D., & Raedeke, T.D. (2008). *Sport psychology for coaches*. HK, USA: Human Kinetics.

Callow, N., Roberts, R., Bringer, J.D., & Langan, E. (2010). Coach education related to the delivery of imagery: two interventions. *The Sport Psychologist, 24,* 277-299. doi:10.1123/tsp.24.3.277

Callow, N., Smith, M. J., Hardy, L., Arthur, C.A., & Hardy, J. (2009). Measurement of transformational leadership and its relationship with team cohesion and performance level. *Journal of Applied Sport Psychology*, *21*, 395-412. doi: 10.1080/10413200903204754

Chelladurai, P., & Saleh, S.D. (1980). Dimensions of leader behavior in sports: Development of a leadership scale. *Journal of Sport Psychology, 2*, 34-45. doi:10.1123/jsp.2.1.34

Côté, J., Yardley, J., Hay, J., Sedgwick, W., & Baker, J. (1999). An exploratory examination of the Coaching Behavior Scale for Sport. *Avante, 5,* 82–92.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: new guidance. Retrieved from https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/

Deci, E.L., & Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227-268. doi: 10.1207/S15327965PLI1104_01

Durand-Bush, N., Salmela, J.H., & Green-Demers, I. (2001). The Ottawa Mental Skills Assessment Tool (OMSAT-3*). *The Sport Psychologist, 15*, 1-19. doi:10.1123/tsp.15.1.1

Edwards, J., Law, B., & Latimer-Cheung, A. (2012). Effects of an imagery workshop on coaches' encouragement of imagery use. *International Journal of Sports Science & Coaching, 7*, 317-331. doi:10.1260/1747-9541.7.2.317

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39-50. doi:10.2307/3151312

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, Florida: Chapman and Hall/CRC Press.

Gould, D., Damarjian, N., & Medbery, R. (1999). An examination of mental skills training in junior tennis coaches. *The Sport Psychologist*, *13*, 127-143. doi:10.1123/tsp.13.2.127

Gould, D., Dieffenbach, K., & Moffett, A. (2002). Psychological characteristics and their development in Olympic champions. *Journal of Applied Sport Psychology*, *14*, 172-204. doi:10.1080/10413200290103482

Hall, C.R., & Rodgers, W.M. (1989). Enhancing coaching effectiveness in figure skating through a mental skills training program. *The Sport Psychologist*, *3*, 142-154. doi:10.1123/tsp.3.2.142

Hall, N., Jedlic, B., Munroe-Chandler, K., & Hall, C. (2007). The effects of an education program on coaches' encouragement of imagery use. *International Journal of Coaching Science, 1,* 79-86. doi: 10.1260/1747-9541.7.2.317

Hanton, S., Mellalieu, S.D., & Hall, R. (2004). Self-confidence and anxiety interpretation: A qualitative investigation. *Psychology of Sport and Exercise 5,* 477-495. doi:10.1016/S1469-0292(03)00040-2

Hardy, L., Roberts, R., Thomas, P.R., & Murphy, S.M. (2010). Test of Performance Strategies (TOPS): Instrument refinement using confirmatory factor analysis. *Psychology of Sport and Exercise, 11,* 27-35. doi:10.1016/j.psychsport.2009.04.007

Harwood, C. (2008). Developmental consulting in a professional football academy: The 5C's coaching efficacy program. *The Sport Psychologist, 22,* 109-133. doi:10.1123/tsp.22.1.109

Harwood, C., Barker, J.B., & Anderson, R. (2015). Psychosocial development in youth soccer players: Assessing the effectiveness of the 5Cs intervention program. *The Sport Psychologist, 29,* 319-334. doi:10.1123/tsp.2014-0161

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modelling: A Multidisciplinary Journal, 6,* 1-55. doi:10.1080/10705519909540118

Jang, H., Reeve, J., & Deci, E.L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology, 102,* 588-600. doi:10.1037/a0019682

Jedlic, B., Hall, N., Munroe-Chandler, K., & Hall, C. (2007). Coaches' encouragement of athletes' imagery use. *Research Quarterly for Exercise & Sport, 78,* 351-363. doi:10.1080/02701367.2007.10599432

Kuhn, D., & Dean, D. Jr. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into Practice, 43*, 268-273. doi:10.1207/s15430421tip4304 4

Liao, C.M., & Masters, R.S.W. (2001). Analogy learning: A means to implicit motor learning. *Journal of Sports Sciences, 19*, 307-319. doi:10.1080/02640410152006081

Lyle, J. (2002). *Sport coaching concepts: A framework for coaches' behaviour*. Oxon, UK. Routledge.

Mageau, G.A., & Vallerand, R.J. (2003). The coach-athlete relationship: A motivational model. *Journal of Sport Sciences, 21*, 883-904. doi:10.1080/0264041031000140374

Markland, D. (2007). The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences, 42*, 851-858. doi:10.1016/j.paid.2006.09.023

Markland, D., & Tobin, V.J. (2010). Need support and behavioural regulations for exercise among exercise referral scheme clients: The mediating role of psychological need satisfaction. *Psychology of Sport and Exercise, 11*, 91-99. doi:10.1016/j.psychsport.2009.07.001

Marshall, C., & Rossman, G.B. (1995). *Designing qualitative research*. Thousand Oaks, CA: Sage.

Muthén, B., & Asparouhov, T. (2012). Bayesian Structural Equation Modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313-335. doi:10.1037/a0026802

Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis. *Mplus Web Notes*, 1-48.

Myers, N.D., Ntoumanis, N., Gunnell, K.E., Gucciardi, D.F., & Lee, S. (2017). A review of some emergent quantitative analyses in sport and exercise psychology. *International Review of Sport and Exercise Psychology, 11,* 77-100. doi:10.1080/1750984X.2017.1317356

Niven, A.G., & Markland, D. (2016). Using self-determination theory to understand motivation for walking: Instrument development and model testing using Bayesian structural equation modelling. *Psychology of Sport and Exercise, 23*, 90-100. doi:10.1016/j.psychsport.2015.11.004

Paquette, K.J., & Sullivan, P. (2012). Canadian curling coaches' use of psychological skills training. *The Sport Psychologist, 26,* 29-42. doi:10.1123/tsp.26.1.29

Patton, M.Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Potrac, P., & Cassidy, T. (2006). The coach as 'more capable other'. In R.L. Jones (Ed.), *The sports coach as educator: re-conceptualising sports coaching* (pp. 39-50). London: Routledge.

Ravizza, K. (2010). Increasing awareness for sport performance. In J.M. Williams (Ed.), *Applied sport psychology: Personal growth to peak performance* (6th ed., pp. 189-200). New York: McGraw-Hill.

Schraw, G., & Dennison, R.S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19,* 460-475. doi:10.1006/ceps.1994.1033

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology, 21,* 277–287. doi:10.1002/acp.1340

Silva, M.N., Vieira, P.N., Coutinho, S.R., Minderico, C.S., Matos, M.G., Sardinha, L.B., & Teixeira, P.J. (2010). Using self-determination theory to promote physical activity and

954 weight control: a randomized controlled trial in women. *Journal of Behavioral*
955 *Medicine, 33,* 110-122. doi:10.1007/s10865-009-9239-y
956 Smith, R.E., Schutz, R.W., Smoll, F.T., & Ptacek, J.T. (1995). Development and validation of
957 a multidimensional measure of sport-specific psychological skills: The Athletic
958 Coping Skills Inventory-28. *Journal of Sport and Exercise Psychology, 17,* 379-398.
959 doi:10.1123/jsep.17.4.379
960 Sparkes, A. C., & Smith, B. (2014). *Qualitative research methods in sport, exercise and*
961 *health: From process to product*. Oxon: Routledge.
962 Sperling, R.A., Howard, B.C., Miller, L.A., & Murphy, C. (2002). Measures of children's
963 knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*,
964 51-79. doi:10.1006/ceps.2001.1091
965 Thelwell, R.C., Greenless, I.A., & Weston, N.J.V. (2006). Using psychological skills training
966 to develop soccer performance. *Journal of Applied Sport Psychology, 18,* 254-270.
967 doi:10.1080/10413200600830323
968 Thomas, P.R., Murphy, S.M., & Hardy, L. (1999). Test of Performance Strategies:
969 development and preliminary validation of a comprehensive measure of athletes'
970 psychological skills. *Journal of Sports Sciences, 17,* 697–711.
971 doi:10.1080/026404199365560
972 Tremayne, P., & Newberry, G. (2005). Mental skill training program for children. In F.
973 Hackfort, J. Duda, & R. Lidor (Eds.), *Handbook of research in applied sport and*
974 *exercise psychology: International Perspectives* (pp.93-106). West Virginia
975 University: Fitness Information Technology.
976 Wagstaff, C.R., Arthur, C.A., & Hardy, L. (2018). The development and initial validation of
977 a measure of coaching behaviors in a sample of army recruits. *Journal of Applied*
978 *Sport Psychology, 30,* 341-357. doi:10.1080/10413200.2017.1384937

Weinberg, R.S., & Comar, W. (1994). The effectiveness of psychological interventions in competitive sport. *Sports Medicine, 18*, 406-18. doi:10.2165/00007256-199418060-00005

Weinberg, R., & Gould, D. (2015). *Foundations of sport and exercise psychology* (5th ed.). Champaign, IL: Human Kinetics Press.

Weinberg, R.S., & Williams, J.M. (2010). Integrating and implementing a psychological skills training program. In J.M. Williams (ed.), *Applied Sport Psychology: personal growth and peak performance* (6th Ed., pp. 361-391). New York, NY: McGraw Hill.

Whitmore, J. (2009). *Coaching for performance: Growing human potential and purpose: The principles and practice of coaching and leadership.* London: Nicolas Brealey International.

Williams, J.M., Kenow, L.J., Jerome, G.J., Rogers, T., Sartain, T.A., & Darland, G. (2003). Factor structure of the Coaching Behavior Questionnaire and its relationship to athlete variables. *The Sport Psychologist, 17*, 16-34. doi:10.1123/tsp.17.1.16

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist, 16*, 75-84.

Vealey, R.S. (1988). Future directions in psychological skill training. *The Sport Psychologist, 2*, 318-336. doi:10.1123/tsp.2.4.318

Table 1

Study 2 Factorial Validity Results for CPS-F and CPS-NS including Fit Statistics, Standardised Factor Loadings and 95% Credibility Intervals

			Difference between observed and replicated χ^2 95% CI	
	BSEM Fit statistics	PP*p*	Lower 2.5%	Upper 2.5%
CPS-F	16-item Non-Informative	.000	282.83	368.56
	12-item Informative Priors (cross- loadings)	.000	23.07	115.72
	11-item Non-informative	.000	36.73	101.54
	11-item Informative Priors (cross-loadings)	.002	16.53	82.65
	11-item Informative priors (cross-loadings + residual correlations)	528	-35.50	34.35
CPS-NS	14-item Non-Informative	.000	241.98	319.06
	9-item Non-Informative	.000	55.66	110.72
	9-item Informative Priors (cross-loadings)	.000	51.83	107.53
	9-item Informative Priors (cross-loadings + residual correlations)	.49	-28.73	29.52
	8-item Non-informative	.006	7.07	56.81
	8-item Informative priors (cross loadings)	.010	4.90	55.90
	8-item Informative priors (cross-loadings + residual correlations)	.51	-27.20	25.87

	Standardised factor loadings for final items	Observation	Targeted Cueing	Instruction
CPS-F	My coach picks up on my use of goal setting.	**.91 [.68,1.14]**	.02 [-.14, .17]	-.01 [-.21,.18]
	My coach notices how much I use goal setting.	**.92 [.70,1.14]**	.001 [-.15,.15]	.01 [-.19,.19]
	My coach observes my use of goal setting.	**.89 [.66,1.13]**	-.003 [-.17,.15]	.03 [-.18,.22]
	My coach includes specific goals in his/her instructions.	-.05 [-.22,.13]	**.87 [.63,1.09]**	-.05 [-.23,.12]
	My coach talks about specific goals to help me be in the right mental state.	.03 [-.15,.20]	**.78 [.54,1.00]**	.04 [-.14,.22]
	My coach describes specific goals to make things easier to understand.	-.008 [-.18,.16]	**.85 [.62,1.05]**	.002 [-.18,.17]
	My coach tells me technical information by talking about specific goals.	-.006 [-.18,.17]	**.86 [.64,1.07]**	-.005 [-.18,.17]
	My coach talks about specific goals to motivate me.	.05 [-.13,.22]	**.79 [.58,1.01]**	.03 [-.15,.20]
	My coach tells me to use goal setting when I'm doing my sport.	-.006 [-.21,.18]	.02 [-.15,.17]	**.90 [.66,1.14]**
	My coach asks me to think about using goal setting when I'm doing my sport.	.04 [-.17,.22]	.02 [-.13,.17]	**.89 [.67,1.12]**
	My coach instructs me to use goal setting	.008 [-.19,.20]	-.02 [-.19,.14]	**.91 [.67,1.14]**

		Explanation Provision	Seeking Athlete Involvement
CPS-NS	My coach makes it clear what to expect from using goal setting.	**.91 [.69,1.12]**	.01 [-.19, .21]
	My coach gives me good advice about goal setting.	**.92 [.71,1.13]**	.02 [-.19,.22]
	My coach explains why using goal setting could help my performance	**.92 [.71,1.13]**	.01 [-.19,.20]
	My coach makes it clear what I need to do to get positive effects from using goal setting.	**.92 [.71,1.13]**	-.004 [-.20,.19]
	My coach encourages me to take my own initiative about using goal setting.	.004 [-.20,.21]	**.84 [.60,1.06]**
	My coach provides me with a range of ways to use goal setting.	.03 [-.18,.23]	**.91 [.70,1.13]**
	My coach and I discuss using goal setting.	-.001 [-.20,.19]	**.90 [.68,1.11]**
	My coach takes into account my needs when speaking with me about goal setting.	.005 [-.19,.20]	**.93 [.72,1.13]**

Note. PP*p* = posterior predictive *p* value; BSEM = Bayesian Structural Equation Modelling. Factor loadings and 95% credibility intervals in bold correspond to the items in each row.

Table 2

The Means, Standard Deviations and Bivariate Correlations between CPS-F, CPS-NS, CBS-S and MAI-S in Study 3.

Scale	Subscale	Mean	*SD*	1	2	3	4	5	6
CPS-F	1. Observation[a]	1.33	1.12	-					
CPS-F	2. Targeted Cueing[a]	2.05	.98	.67**	-				
CPS-F	3. Instruction[a]	1.58	1.17	.83**	.70**	-			
CPS-NS	4. Explanations of PS[a]	1.75	1.16	.73**	.68**	.79**	-		
CPS-NS	5. Seeking Athlete Involvement[a]	1.58	1.06	.79**	.67**	.81**	.90**	-	
CBS-S	6. Mental Preparation[b]	4.93	1.40	.48**	.45**	.51**	.55**	.52**	-
MAI-S	7. Awareness of PS[c]	64.41	16.14	.24**	.24**	.29**	.33**	.34**	.21**

Note. ** correlation is significant $p < .01$. CPS-F = Coaching of Psychological Skills Scale – Fundamental; CPS-NS = Coaching of Psychological Skills Scale – Need Support; CBS-S = Coaching Behavior Scale for Sport; MAI-S = Mental Awareness Inventory for Sport.
[a]variable rated on a 0-4 scale.
[b]variable rated on a 1-7 scale.
[c]variable rated on a 1-100 scale.

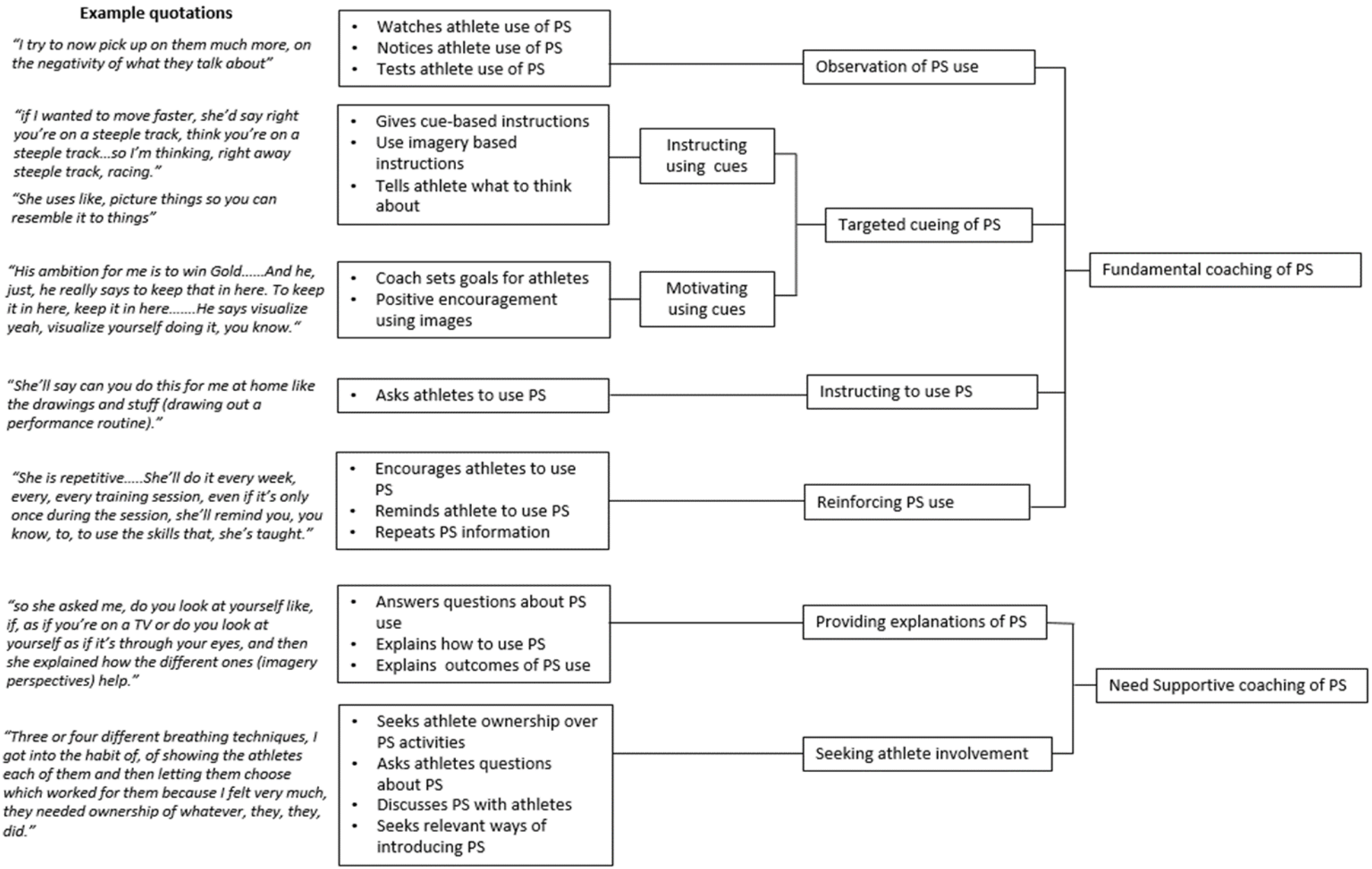

Figure 1. Results of Study 1. A hierarchical content tree of the coaching of psychological skills (PS) and example quotations