

**University of Birmingham**

---

**From the Selected Works of Jennifer Cumming**

---

2012

# Assessing Mental Skill and Technique Use in Applied Interventions: Recognizing and Minimizing Threats to the Psychometric Properties of the TOPS

Charlottee Woodcock, *University of Birmingham*

Joan L. Duda, *University of Birmingham*

Jennifer Cumming, *University of Birmingham*

Lee-Ann Sharp

Mark J. G. Holland, *University of Birmingham*



Available at: [https://works.bepress.com/jennifer\\_cumming/22/](https://works.bepress.com/jennifer_cumming/22/)

# Assessing Mental Skill and Technique Use in Applied Interventions: Recognizing and Minimizing Threats to the Psychometric Properties of the TOPS

**Charlotte Woodcock, Joan L. Duda, and Jennifer Cumming**  
University of Birmingham

**Lee-Ann Sharp**  
University College Plymouth St. Mark & St. John

**Mark J.G. Holland**  
University of Birmingham

Drawing from the experiences of the authors in developing, conducting, and evaluating sport psychology interventions, several considerations are highlighted and recommendations offered for effective psychometric assessment. Using the Test of Performance Strategies (TOPS; Thomas, Murphy, & Hardy, 1999) as a working example, opportunities for bias to undermine a measure's validity and reliability are discussed with reference to a respondent's four cognitive processes: (a) comprehension, (b) retrieval, (c) decision-making, and (d) response generation. Further threats to an instrument's psychometric properties are highlighted in the form of demand characteristics athletes perceive in the environment. With these concerns in mind, several recommendations are made relating to the process of questionnaire administration and how possible compromises to the psychometric soundness of measures used in applied interventions can be minimized.

The credibility of mental skills training (MST) within applied practice is dependent on each practitioner's ability to effectively evaluate the intended intervention outcomes (Vealey, 1988). For multimodal MST programs in which athletes are taught techniques (e.g., goal setting, imagery, relaxation, and self-talk) for improvement in their mental skills (such as attentional or arousal control), it is important to establish at the very least whether athletes are using the techniques taught (Anderson, Mile, Mahoney, & Robinson, 2002). It is otherwise difficult to attribute change in desirable outcomes to a MST intervention.

---

Woodcock, Duda, Cumming, and Holland are with the School of Sport and Exercise Sciences, University of Birmingham, Edgbaston, UK. Sharp is with the University College Plymouth St. Mark & St. John, Plymouth, UK.

The Test of Performance Strategies (TOPS; Thomas, Murphy, & Hardy, 1999) has been the assessment of choice when examining the frequency in which athletes employ specific mental skills and techniques (e.g., Fletcher & Hanton, 2001). The scale is best thought of as two separate measures, one contextualized for competition and the other for practice. Sixteen subscales, eight for each of the sport settings, cover a variety of mental skills and techniques underlying athletic performance. The mental techniques tapped include goal setting, imagery, relaxation, and self-talk; while the mental skills targeted are automaticity, arousal control, emotional control, and attentional control in practice or negative thinking in competition.

The TOPS has often been adopted in applied research to assess intervention effectiveness. For example, Cohen, Tenenbaum, and English (2006) administered the TOPS before and after an emotion regulation program to ascertain intervention effectiveness in increasing athletes' use of targeted psychological skills and techniques. Similarly, Kirschenbaum, Owens, and O'Connor (1998) used information gleaned from pre- and postintervention TOPS subscale scores to conclude athletes' use of self-talk and emotional control had become more frequent following an applied intervention.

Psychometric concerns surrounding the use of questionnaires such as the TOPS have been an on-going issue in the field of sport psychology where self-report measures dominate assessment procedures (Conroy, Kaye, & Schantz, 2008). For these tools to be effective in evaluating the effectiveness of applied interventions, Beckmann and Kellmann (2003) outline criteria to be considered by the applied practitioner. Specifically, they argue that an instrument must be: (a) reliable and valid for the population in question, (b) perceived to be useful, and (c) answered honestly. Several phases of assessment were consequently proposed by Beckmann and Kellmann to satisfy these criteria.

When selecting a test, the sport psychologist must consider the psychometric properties of validity and reliability (Beckmann & Kellmann, 2003). A valid questionnaire measures the construct it purports to measure. During instrument development, several types of validity are often addressed. Beginning with the wording of individual items, face validity suggests items capture some or all aspects of a particular factor or construct. If several items, intended to represent a single factor, converge together, then a measure is considered to have convergent validity. At this factor or subscale level, concurrent and discriminant validity is assessed. Concurrent validity refers to positive correlations between a factor and theoretically related constructs. Conversely, but equally important, discriminant validity assures a factor is distinct from theoretically unrelated constructs. Finally, predictive validity demonstrates the assessed construct's ability to behave in a way it is theoretically intended (either as an antecedent or consequence) in relation to other variables.

For a questionnaire to be valid, it must be reliable. The reliability of an instrument reflects the consistency with which it measures constructs. That is to say, a reliable questionnaire is one that is trustworthy and relatively free of error (Tenenbaum, Kamata, & Hayashi, 2007). Internal reliability, often indicated by Cronbach's  $\alpha$ , reflects consistency between items intended to measure the same factor. Alpha values are usually reported in studies employing self-report measures (Haggar & Chatzisarantis, 2009), with a cut-off value of .70 deemed to be acceptable.

Another type of reliability that is particularly important for the evaluation of sport psychology interventions is test-retest reliability. Often assumed to be evident in longitudinal research, test-retest reliability reflects a questionnaire's consistency

in measuring a construct over time. If a measure exhibits this type of reliability, then observed changes between pre- and postintervention scores are likely to indicate changes in the measured construct as a function of the delivered intervention. In contrast, a measure low in test-retest reliability suggests that changes in questionnaire scores may be due to unknown measurement error rather than the impact of the intervention per se (Schutz, 1998).

Indicators of measurement validity and reliability are offered in applied research (e.g., Cohen et al., 2006; Woodcock, Holland, Sharp, Duda, & Cumming, 2010a), however applied studies rarely consider the impact the process of questionnaire administration has on these psychometric factors. Indeed, Beckmann and Kellmann (2003) stress the need for additional considerations to be made during measure administration to protect against potential biases within the applied setting. To guard against social desirability (Huang, Liao, & Chang, 1998) for example, it is important to help athletes understand the purpose of the assessment and provide them with accurate information on who will have access to their responses. Building a level of confidence and trust with athletes is deemed critical for facilitating open and honest responses. Further, the timing of questionnaire administration and the environment within which responses are generated should be considered to promote maximal athlete engagement. Athletes will be more receptive to completing questionnaires when they are in an undisturbed and relaxing environment. In such circumstances, they are more likely to devote their attention to the task. By comparison, questionnaires presented immediately after training when athletes are tired and hungry, or in the minutes before an important competition, are unlikely to be well received or responded to with concerted focus.

The intended audience for Beckmann and Kellmann's (2003) thoughtful piece appears to be the exclusively applied consultant rather than the applied practitioner-researcher. According to Beckmann and Kellmann, the purpose of their assessment was to inform an athlete's needs analysis, as well as supporting coach decisions. For example, the physical and psychological demands placed on athletes by a coach may be revised if questionnaire responses suggest an athlete's experiencing symptoms of overtraining. In contrast, when conducting applied research in the field, the purpose of a questionnaire is to monitor changes in the assessed variables over time rather than directly impacting individual behaviors and applied decisions made by support staff. Thus, when administering psychological assessments in an applied research context, we have found the considerations proffered by Beckman and Kellmann only partially beneficial (e.g., Holland, Sharp, Woodcock, Cumming, & Duda, 2010a). We faced further issues when administering psychometric assessments for research goals, rather than for individual athlete assessment.

Given it is in applied research where theory and practice more closely dovetail together, and credibility of the field rests, it is critical that the assessment procedures employed do not compromise the validity and reliability of a targeted measure. Consequently, a thorough deliberation of the process of questionnaire administration in applied research settings (to minimize potential sources of bias) is paramount to obtain data that are valid and reliable. In the interests of obtaining quality data from which appropriate inferences can be made, the aim of this paper is to share lessons learned from our applied research experiences using the TOPS as a working example. In so doing, we aim to extend Beckmann and Kellmann's (2003) recommendations for employing psychometric assessment to evaluate applied interventions.

## Where Lessons Were Learned: Applied Research Experiences

### MST Program

To illustrate the proposed considerations for the use of psychological assessments in applied research, the present paper draws from two “real world” studies. The first study involved a season-long MST program designed for and delivered to adolescent elite male rugby union players in the UK (Woodcock, Holland, Sharp, Duda, & Cumming, 2010b). The project included a season-long pilot program followed by a revised intervention phase delivered the following season. Fifty-seven players ( $M_{age} = 15.25$  years,  $SD = 0.38$ ) participated in the pilot program with a matched control group of 46 players ( $M_{age} = 15.43$  years,  $SD = 0.26$ ). One hundred and six players ( $M_{age} = 15.41$  years,  $SD = 0.24$ ) received the revised intervention, which had a matched control group involving 54 athletes ( $M_{age} = 15.38$  years,  $SD = 0.28$ ).

Psychometric measures were administered pre- and postintervention to test specific hypotheses and triangulate qualitative data. Psychological variables assessed included symptoms of athlete burnout, motivation regulations, competitive anxiety, and self-confidence. Furthermore, frequency of psychological skill and technique use was examined to determine whether the MST program was effective in increasing the employment of athletes’ mental techniques and development of mental skills.

### Emotion Regulation Program

The second example stems from lessons learned in a multiple case study. Two golfers and three tennis players of varying ages ( $M_{age} = 18.33$  years,  $SD = 3.92$ ) and levels of expertise followed an individualized 8-week emotion regulation program. Similar to the MST program, psychometric assessments were administered to measure psychological variables of interest such as general affective states experienced in sport. Again, frequency of mental skill and technique use was determined to evaluate the intervention effectiveness in increasing athlete employment of these desired behaviors. In both studies, the TOPS (Thomas et al., 1999) ascertained frequency of mental skill and technique use.

Our focus in this paper is not on research design or the content of the interventions per se. Instead, we discuss the administration considerations made for the TOPS with an overall aim to gather data that were both valid and reliable. In essence, the overarching aspiration is to more accurately reflect athletes’ frequency of mental skill and technique use pre- and postintervention. With such issues in mind, we examine potential biases that threaten the validity and reliability of the TOPS. Each opportunity for bias is outlined in turn and, in making our points, we pull from recommendations made in the literature as well as our recent work on implementing a MST program in a team sport setting and case studies centered on promoting athletes’ emotion regulation. Drawing from these examples of lessons learned, recommendations are made for minimizing compromises to sound measurement in applied research.

### Threats to the Psychometric Soundness of the TOPS

Self-report measures are open to being influenced by a number of biases tied to cognitive processes of the respondent as well as demand characteristics present in

the environment (Brener, Billy, & Grady, 2003). Following examination of a questionnaire's psychometric properties, considerations need to be made to ensure the questionnaire's validity and reliability is not undermined during its administration.

Drawing from several cognitive models, Jobe (2000) described four key processes that influence self-report responses: (a) item comprehension and interpretation, (b) information retrieval, (c) decision-making, and (d) response generation. It has been hypothesized that response errors can arise during each phase of the cognitive process (e.g., Brener et al., 2003; Jobe, 2000). Comprehension relates to how an athlete interprets the meaning of an item. Retrieval cues generated from this comprehension process are used to search for relevant self-schema or information about the self relating to the item (Markus, 1977). The adequacy of this retrieved information in terms of participants' interpretation and understanding of the item is evaluated during the decision-making phase. If the retrieved information is deemed favorable to offer an acceptable answer, then response generation follows. Opportunities for bias in athletes' responses to psychological measures are described in turn below.

## Item Comprehension and Interpretation

Issues of comprehension occur when ambiguous or unfamiliar language is used (Brener et al., 2003), and threaten reliability if an athlete does not understand an item well enough to offer a response. Alternatively, validity is undermined when items are seemingly understood but then interpreted differently to researchers' intentions. Hence recommendations to use simple everyday language in our measurement tools have been proposed (e.g., Choi & Pak, 2005).

To measure mental skill and technique use, items that comprise the TOPS subscales include common nomenclature phrases and concepts from sport psychology such as "goal setting" and "self-talk". In previous work, adolescent athletes (13–15 years) have articulated a level of mental skill knowledge consistent with accepted definitions in the literature (McCarthy, Jones, Harwood, & Olivier, 2010). Nevertheless, it appears that this level of knowledge can be progressed further. For example, following participation in MST programs, senior and national level athletes reported an increase in their understanding of such psychological skills and techniques (Brewer & Shillinglaw, 1992). These changes in athletes' knowledge suggest the literal meaning of terms, such as goal setting, are likely to alter following their participation in a MST program.

**MST Program.** Item comprehension and interpretation was inferred from players in our MST program. 20 players participated in three focus groups and one interview (see Sharp, Woodcock, Holland, Duda, & Cumming, 2010a), and were asked about their knowledge of MST at the beginning of the season when preintervention measures were administered. One player commented that "the majority of the stuff I had not done before" and another player expressed his initial desire to "know all the proper...terminology". These reports suggest a limited preintervention understanding of sport psychology terms responded to in TOPS items. Thus an educational MST program aimed to enhance athletes' knowledge is likely to alter an individual's comprehension and interpretation of questionnaire items.

## Information Retrieval

Following comprehension and interpretation of questionnaire items, a respondent searches for relevant self-schema, or self-knowledge stored in long-term memory

from which to formulate a response (Markus, 1977). Self-knowledge, however, demands a level of self-awareness (Baumeister, 2010), and the latter is also a quality often developed via MST. Vealey (1988) identifies self-awareness as a foundation skill, and is a necessary precursor to efforts of self-control and self-regulation (Duda, Cumming, & Balaguer, 2005; Ravizza, 2006). Changing athletes' level of awareness via MST consequently impacts what *self-knowledge* is available when responding to questionnaire items pre- to postintervention. This variability in self-awareness over time can provide an extraneous effect on observed responses to TOPS items when evaluating applied programs.

**MST Program.** Changes in athlete awareness and self-knowledge of mental skill and technique use were indicated following the MST program. A young rugby player emphasized how the MST program “made me aware of what I was doing”. For one athlete this increase in awareness led to a realization that “I’ve always instinctively set goals during games...it’s [the MST program] just made me more aware of it”. In this player’s case, the retrieval of self-schema relating to goal setting was more pronounced postintervention. This athlete’s proposed increases in awareness suggest his responses to items on the TOPS may initially have been misrepresented because relevant information was not available to conscious retrieval processes.

## Decision Making

The self-schema retrieved from memory is reviewed during the decision making phase. The degree to which an athlete’s response to an item reflects this self-schema depends on the motivational processes of the participant. Individuals may choose to ignore valid self-referent information in favor of what they believe to be more acceptable accounts to avoid criticism or gain approval from others (Huang et al., 1998). An attempt by athletes to present themselves in socially acceptable ways threatens the validity of self-report measures. As a consequence this bias may mask relationships between independent and dependent variables as a consequence (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Several demand characteristics that are likely to influence the cognitive processes of decision making and foster socially desirable responses from athletes include perceived coach evaluation, athlete expectations of intervention outcome, and team interactions.

If athletes perceive psychometric instruments as a form of personal evaluation, the social desirability factor of impression management can come into force. This is more likely to be marked when athletes believe coaches have access to their responses. Furthermore, if an athlete harbors concerns about playing time in a match or even for team selection, this effect can be further confounded. In such contexts, athletes are likely to offer more flattering representations of themselves during the decision making phase of item response, than actual self-schema suggest. In a study with swimmers, Young and Starkes (2006) acknowledged that athletes may have presented themselves in an overly favorable light due to such social desirability pressures.

Athletes may also respond in a way that is consistent with perceived intervention outcomes. A similar behavior is observed in clinical psychology where practitioners observe the “hello-goodbye effect” (Gravetter & Forzano, 2008). In the first instance, patients seeking treatment will exaggerate symptoms to ensure referral for treatment and minimize them following an intervention to reinforce any progress they have made. Although we do not perceive ourselves to be “treating

patients” in applied sport psychology educational programs, athletes who engage with the MST process may expect positive change in mental skill and technique use, and provide responses to reflect these anticipated gains.

A further demand characteristic potentially influencing athletes’ decision making for item response is knowledge of peer responses. Beckmann and Kellmann (2003) had concerns over a rowing team’s jocular interactions during questionnaire completion. A particular worry, in the case of this rowing team, was that some athletes would assume another’s responses to be correct and copy them for their own. Unfortunately, ideal conditions for a relaxed and comfortable environment where athletes are not disturbed and are unable to readily see teammates’ responses are rarely found in a team environment.

**MST Program.** In our applied research concerning the effectiveness of the implemented MST, similar demand characteristics were observed. These young talented players were under constant evaluation from coaching staff, who were faced with the task of reducing the squad by half over the course of the season. Thus, we had concerns that athletes misinterpreted the purpose of the psychological assessments. Due to the pressures of selection, we feared athletes perceived questionnaires as part of this squad determination process, although it was explicitly communicated that MST sessions played no role in selection and any data collected from the program were confidential.

Furthermore, social validation interviews suggested athletes held positive expectations for the intervention outcome. Players reported entering the program wanting “more control over ...emotions”, “[to] learn new things”, and “to improve rugby”. A player wishing to fulfill these aspirations for self-improvement is likely to exaggerate responses in line with the expected impact of the intervention.

Similar to Beckmann and Kellmann’s (2003) experience with a rowing team, data collection for the evaluation of the MST program was conducted within a team environment. We observed players exchanging words with fellow players while completing the questionnaire, and feared they might be more concerned about giving a right answer instead of responding open and honestly based on their individual rugby experiences.

## Response Generation

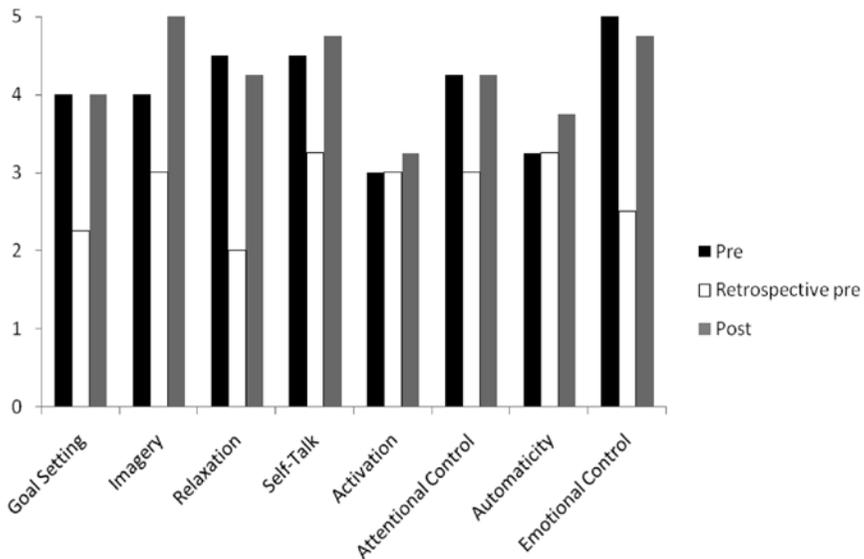
The final cognitive process pertinent to response variability is to match what is considered to be an appropriate response to the available scale (Morsbach & Prinz, 2006). In psychology studies, items are often responded to on a Likert scale ranging from 1 to 5 or 1 to 7. What number is selected will depend on factors already discussed in previous phases, as well as the range of options available on a given scale. If an athlete at the first time point rates him or herself as a 4 (often) or 5 (always) on a 5-point scale, such as for the response options available on the TOPS, there exists a small margin for improvement (if any) to be captured at subsequent time points. This scale response limitation is known as the ceiling effect, and has been previously reported in sport psychology intervention research (e.g., Filby, Maynard, & Graydon, 1999).

**Emotion Regulation Program.** A ceiling effect was observed between pre- and postintervention TOPS subscale scores for the emotion regulation program. Namely, a 22 year-old male golfer reported high ratings on each subscale of the TOPS that

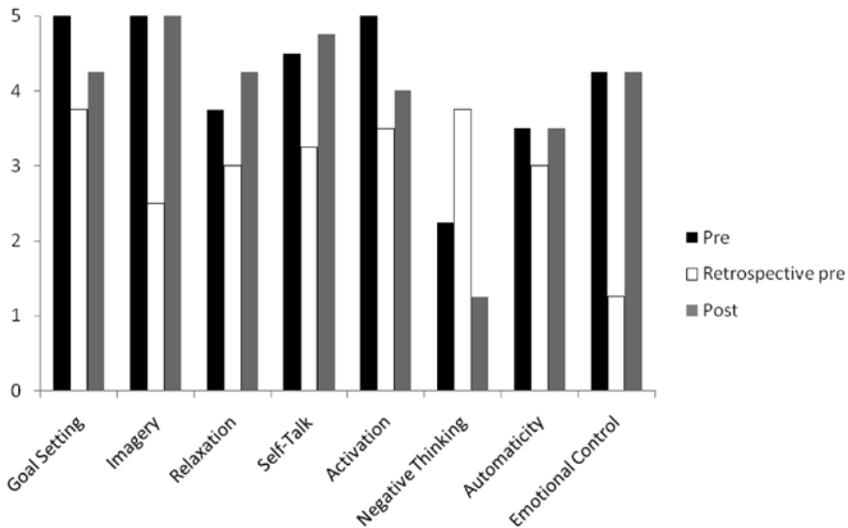
were similar pre- and postintervention, suggesting no changes in the frequency of mental skill and technique use had occurred. To confirm whether this lack of change was due to a ceiling effect, the TOPS was administered a third time. On this occasion the athlete recalled the frequency of his mental skill and technique use before participating in the implemented emotion regulation program. This reflective measurement provided a retrospective preintervention score for each TOPS subscale. When retrospective preintervention ratings were compared with postintervention scores, an increase in frequency across all TOPS subscales in both practice and competition settings was observed, with the exception of negative thinking where scores decreased (see Figure 1 and 2).

## Minimizing Threats to the Psychometric Soundness of the TOPS

A number of threats to the validity and reliability of the TOPS have been outlined in this present paper, including biases in athletes' cognitive processing of items, as well as the potential impact of demand characteristics on response generation. These identified risks to the psychometric properties of the TOPS led us to implement a number of changes in how the questionnaire was administered in a revised MST program. These methodological revisions are presented below to provide guidance to other researchers as to how threats to the psychometric soundness of the TOPS may be minimized.



**Figure 1** — Practice TOPS subscale scores for a 22 year old golfer pre- and postintervention following an 8-week emotion regulation program



**Figure 2** — Competition TOPS subscale scores for a 22 year old golfer pre- and postintervention following an 8-week emotion regulation program

## Item Comprehension and Interpretation

In an attempt to overcome misinterpretation of items or potential changes in athletes' understanding of mental skills terminology from pre- and postintervention, we verbally defined terms targeted within the TOPS subscale items. Goal setting and self-talk were identified to be terms that had marked potential for misinterpretation by our participants. Thomas and colleagues (1999) do not offer definitions for these mental techniques. Therefore, we developed an explanation for each term that stemmed from the literature and best reflected each construct for the participant age group (e.g., McCarthy et al., 2010). Goal setting was defined as "the deliberate setting of specific performance targets to be achieved in your sport". Similarly, a definition of self-talk was offered to athletes that stated "self-talk refers to the content of your thoughts, or the dialogue you have with yourself, when participating in your sport". These simplified definitions were favored over the more precise conceptual distinctions cited in research papers (e.g., Hardy, 2006) as they were deemed more suitable for adolescent athletes. We provided the definitions pre- and postintervention and encouraged the athletes to keep these in mind when responding to items. In doing so, we hoped to reduce sources of nonrandom error such as variability in knowledge or awareness of mental skills.

Questionnaire administrators were also available to athletes involved in the revised MST program to answer any uncertainties concerning item comprehension. Schober, Conrad, and Fricker (2004) found interviewers who responded to requests for clarification, and offered help to participants who required assistance, received more accurate responses. Consequently, administrators periodically circled the room and offered guidance to the athletes who appeared to be having difficulties.

It is worthy to note that young athletes may not always seek administrator assistance and instead favor soliciting guidance from a teammate. Putting a hand up and asking for help in front of the whole squad may be embarrassing for youth athletes and instigate undesired attention from peers. Although we advocate that athletes seek assistance from trained administrators, we felt it better they ask a teammate rather than not ask for any clarification.

Having an administrator present for administration (rather than collecting data online) reduced the risk of having missing data due to athlete uncertainty in item meaning. After athletes returned “completed” inventories, we found administrators were able to spend 5–10 s checking the questionnaires for missing items. Any incomplete measures were immediately returned with the nonresponded items highlighted for subsequent completion by the athlete in question. If athletes had missed an item due to difficulties with its comprehension, an administrator was on hand to help clarify any concerns. This brief method of quality control helped minimize nonresponse and thus allowed us to have more data to analyze statistically (without having to employ different strategies for handling missing data).

## **Information Retrieval**

Giving athletes a clear definition of psychological terms, such as goal setting, will help direct retrieval of self-knowledge that is relevant to make a valid response. Nevertheless it should be noted that access to this information is likely to increase following an MST intervention due to enhanced levels of athlete knowledge of MST skills and techniques and greater awareness of their own behaviors. Although this matter was not specifically addressed in the revised MST program, the lesson learned from the emotion regulation program offers one way to overcome this issue. Collecting retrospective preintervention scores after program completion allows an athlete to rate preintervention behaviors with postintervention awareness and self-knowledge.

## **Decision Making and Response Generation**

Deciding whether retrieved information is acceptable to make a response on a Likert scale, and what that response should be, can be influenced by social desirability. For the revised MST program, it was explained to coaches what variables would be measured and how the data obtained from the athletes would be treated confidentially and analyzed at a group rather than individual level. It was also stressed that responses would be used for research purposes only and would not be available to influence decisions for selection. Communicating the aims of assessment to coaches helped secure support for questionnaire administration. Further, obtaining coach backing promotes athlete engagement in questionnaire completion (Beckmann & Kellmann, 2003).

A similar explanation was given to the athletes participating in the revised MST program. Furthermore, athletes were reassured that no right or wrong answers exist to encourage honest responses (Beckmann & Kellmann, 2003). That is, both the purpose of assessment and confidentiality of item responses were emphasized to encourage player acceptance of retrieved self-schema that consequently informs decision-making and response generation processes.

To further encourage more honest responses to questionnaire items, Beckmann and Kellmann (2003) recommend practitioners develop a good rapport with athletes before questionnaire administration. In intervention research, however, there is typically limited contact time between athlete and researcher to develop a trusting relationship before baseline data are collected. In contrast, postintervention measures are administered after a working alliance has been established. Given the potential mediating role rapport has on item response, it is important to separate the research and applied components of an investigation. Consequently, in our revised MST program, we recruited and trained independent administrators not involved in the delivery of program content to collect questionnaire data at all time points.

## Future Considerations

Enthusiasm to proceed with applied research, before careful consideration of the process involved in questionnaire administration, risks compromising ethical and professional integrity as a researcher and applied practitioner (Vealey & Garner-Holman, 1998). Taking consideration of a questionnaire's validity and reliability for the participant group in question is only the beginning. Threats to the psychometric soundness of a previously established valid and reliable measure when employed in intervention studies also require consideration. In proposing recommendations for minimizing these risks, the present paper aims to extend previous recommendations made by Beckmann and Kellmann (2003). Specifically, strategies are offered to facilitate the valid and reliable assessment of the impact psychological interventions have on athletes' mental skill and technique use in group applied research settings.

The present paper draws extensively from the applied research experiences of the authors in evaluating the frequency of athletes' psychological skill and technique use as measured by the TOPS. The points and proposals made should not be considered as exhaustive. The recommendations proposed stem from lessons learned across several intervention studies, and should be regarded as one of many stepping stones toward obtaining reliable and valid data in applied research settings. It is our hope that these recommendations help researchers make informed decisions in questionnaire selection including the TOPS, its refined successors (e.g., Hardy, Roberts, Thomas, & Murphy, 2010), and beyond, as well as encouraging careful consideration for how such measures are administered. Nevertheless, further research regarding questionnaire administration in minimizing threats to psychometric properties in applied research is required.

Future work is warranted on participant cognitive processes that result in variability in item response. For example, how factors such as item understanding and interpretation can result in error of observed measurement scores is one important area in need of further consideration. A methodological tool for examining these potential sources of response bias is the cognitive interview (Willis, 2005). Recommended for use in questionnaire development, cognitive interviews can offer insight into cognitive phases of item response generation (e.g., Dietrich & Ethrenspiel, 2010; Morsbach & Prinz, 2006).

One method of cognitive interviewing is the think-aloud protocol whereby participants verbally report thought processes as they respond to questionnaire items (Dietrich & Ethrenspiel, 2010). This method has highlighted issues in item

interpretation for young athletes responding to Raedeke and Smith's (2001) Athlete Burnout Questionnaire (Sharp, Woodcock, Holland, Duda, & Cumming, 2010b), and the Lonsdale, Hodge, and Rose's (2008) Behavioral Regulations in Sport Questionnaire (Holland, Sharp, Woodcock, Cumming, & Duda, 2010b). Interested readers are directed to Dietrich and Ethrenspiel (2010) who provide a working example for best practice of think-aloud protocols in the sport sciences.

In addition, future methodological studies should consider the information retrieval and decision-making processes athletes engage in when responding to questionnaire items. In clinical psychology, retrieval cues are used to aid access to self-knowledge (Marsbach & Prinz, 2006). When responding to the TOPS such cues could be used to facilitate more accurate recall. For example, athletes could be asked to respond over a specific time period, such as how frequently a relevant action has occurred over the previous 4 weeks when participating in sport. Further, to enhance the decision-making process, a confidence rating scale may be included. This additional measure would allow athletes to indicate how certain a response rating accurately represents a more objective reality.

The potential for ceiling effects that can mask intervention change was highlighted in the emotion regulation program. One recommendation to guard against this effect is to collect retrospective preintervention scores. However retrospective assessments invite a number of criticisms such as delayed recall and further risks for social desirability in fulfilling expectations for intervention outcome (Young & Starkes, 2006). A potential avenue to overcome these issues is to adopt a change version of the TOPS postintervention. Instead of collecting pre- and postintervention data on a single scale, athletes are asked to compare the frequency with which psychological skill and techniques are employed with initial baseline levels. Although this approach invites biases of social desirability and delayed recall, it has been previously used to overcome ceiling effects in assessing the efficacy of different intervention treatments in clinical settings (e.g., Howorka, Pumprla, Schlusche, Wagner-Nosiska, Schabmann, & Bradley, 2001).

Finally, further examination is needed on how the demands of the assessment environment lead to potential socially desirable responses. More systematic research into athlete perceptions of these demands is required to better inform ways to overcome their influence in eliciting response bias. Such information will help researchers optimize the data collection process to obtain more valid and reliable data when assessing the impact of MST programs.

## Concluding Remarks

It is our hope that the present considerations have highlighted how measurement errors may arise in applied work, and evoke critical thinking in keeping such sources of error to a minimum. Those engaged in applied research should strive to select measures with sound psychometric properties, and minimize potential threats to validity and reliability which are tied to the process of questionnaire administration. These key principles should not be assumed, and researchers need to demonstrate a conscientious approach toward "controlling the controllables" and increasing the quality of data when evaluating applied practice.

## References

- Anderson, A.G., Miles, A., Mahoney, C., & Robinson, P. (2002). Evaluating the effectiveness of applied sport psychology practice: Making the case for the case study approach. *The Sport Psychologist, 16*, 432–453.
- Baumeister, R.F. (2010). The self. In R.F. Baumeister & E.J. Finkel (Eds.), *Advanced social psychology: The state of science* (pp. 139–175). New York, NY: Oxford University Press.
- Beckmann, J., & Kellmann, M. (2003). Procedures and principles of sport psychological assessment. *The Sport Psychologist, 17*, 338–350.
- Brener, N.D., Billy, J.O.G., & Grady, W.R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *The Journal of Adolescent Health, 33*, 436–457.
- Brewer, B.W., & Shillinglaw, R. (1992). Evaluation of a psychological skills training workshop for male intercollegiate lacrosse players. *The Sport Psychologist, 6*, 139–147.
- Choi, B.C.K., & Pak, A.W.P. (2005). A catalogue of biases in questionnaires. *Preventing Chronic Disease, 2*, 1–13.
- Cohen, A.B., Tenenbaum, G., & English, R.W. (2006). Emotions and golf performance: An IZOF-based applied sport psychology case study. *Behavior Modification, 30*, 259–280.
- Conroy, D.E., Kaye, M.P., & Schantz, L.H. (2008). Quantitative research methodology. In T.S. Horn (Ed.), *Advances in sport psychology* (3rd ed., pp. 15–30). Leeds: Human Kinetics.
- Dietrich, H., & Ehrlenspiel, F. (2010). Cognitive interviewing: A qualitative tool for improving questionnaires in sport science. *Measurement in Physical Education and Exercise Science, 14*, 51–60.
- Duda, J.L., Cumming, J., & Balaguer, I. (2005). Enhancing athletes' self regulation, task involvement, and self determination via psychological skills training. In D. Hackfort, J.L. Duda, & R. Lidor (Eds.), *Handbook of applied sport psychology research* (pp. 143–165). Morgantown, WV: Fitness Information Technology.
- Filby, W.C.D., Maynard, I.W., & Graydon, J.K. (1999). The effect of multiple-goal strategies on performance outcomes in training and competition. *Journal of Applied Sport Psychology, 11*, 230–246.
- Fletcher, D., & Hanton, S. (2001). The relationship between psychological skills usage and competitive anxiety responses. *Psychology of Sport and Exercise, 2*, 89–101.
- Gravetter, F.J., & Forzano, L.B. (2008). *Research methods for the behavioral sciences* (Vol. 3, 3rd ed.). Belmont, CA: Wadsworth Cengage Learning.
- Hagger, M.S., & Chatzisarantis, N.L.D. (2009). Assumptions in research in sport and exercise psychology. *Psychology of Sport and Exercise, 10*, 511–519.
- Hardy, J. (2006). Speaking clearly: A critical review of the self-talk literature. *Psychology of Sport and Exercise, 7*, 81–97.
- 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.
- Holland, M.J.G., Sharp, L., Woodcock, C., Cumming, J., & Duda, J.L. (2010a, October). Evaluating MST research procedures: Observations, reflections, and progressions. In J. L. Duda (Discussant) *An action research approach to mental skills training: Experience, evaluation and evolution*. Symposium conducted at the meeting of the Association for Applied Sport Psychology, Providence, RI.
- Holland, M.J.G., Sharp, L., Woodcock, C., Cumming, J., & Duda, J.L. (2010b). Validity and reliability of the Behavioral Regulation in Sport Questionnaire (BRSQ) with youth athletes. *Journal of Sport & Exercise Psychology, 32*, s175–s176.
- Howorka, K., Pumprla, J., Schlusche, C., Wagner-Nosiska, D., Schabmann, A., & Bradley, C. (2000). Dealing with ceiling baseline treatment satisfaction level in patients with diabetes under flexible, functional insulin treatment: Assessment of improvements in treatment satisfaction with a new insulin analogue. *Quality of Life Research, 9*, 915–930.

- Huang, C., Liao, H., & Chang, S. (1998). Social desirability and the Clinical Self-Report Inventory: methodological reconsideration. *Journal of Clinical Psychology, 54*, 517–528.
- Jobe, J.B. (2000). Cognitive processes in self-report. In A. Stone, C.A. Bachrach, J.S. Turkkan, J.B. Jobe, H.S. Kurtzman, & V.S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 25–28). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kirschenbaum, D.S., Owens, D., & O'Connor, E.A. (1998). Smart golf: Preliminary evaluation of a simple, yet comprehensive, approach to improving and scoring the mental game. *The Sport Psychologist, 12*, 271–282.
- Lonsdale, C., Hodge, K., & Rose, E.A. (2008). The Behavioural Regulation in Sport Questionnaire (BRSQ): Instrument development and initial validity evidence. *Journal of Sport & Exercise Psychology, 30*, 323–355.
- Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35*, 63–78.
- McCarthy, P.J., Jones, M.V., Harwood, C.G., & Olivier, S. (2010). What do athletes implicitly understand about psychological skills? *Journal of Clinical Sports Psychology, 4*, 158–172.
- Morsbach, S.K., & Prinz, R.J. (2006). Understanding and improving the validity of self-report of parenting. *Clinical Child and Family Psychology Review, 9*, 1–21.
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Sport Psychology, 89*, 879–903.
- Ravizza, K. (2006). Increasing awareness for sport performance. In J.M. Williams (Ed.), *Applied sport psychology: Personal growth to peak performance* (5th ed., pp. 228–239). New York: McGraw-Hill.
- Raedeke, T.D., & Smith, A.L. (2001). Development and preliminary validation of an athlete burnout measure. *Journal of Sport & Exercise Psychology, 23*, 281–306.
- Schober, M.F., Conrad, F.G., & Fricker, S.S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology, 18*, 169–188.
- Schutz, R.W. (1998). Assessing the stability of psychological traits and measures. In J.L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 393–408). Morgantown, WV: Fitness Information Technology.
- Sharp, L., Woodcock, C., Holland, M.J.G., Duda, J.L., & Cumming, J. (2010a, October). Youth athletes responses to a mental skills training program. In J. L. Duda (Discusant), *An action research approach to mental skills training: Experience, evaluation, and evolution*. Symposium conducted at the meeting of the Association for Applied Sport Psychology, Providence, RI.
- Sharp, L., Woodcock, C., Holland, M.J.G., Duda, J.L., & Cumming, J. (2010b). Validation of the Athlete Burnout Questionnaire with youth athletes. *Journal of Sport & Exercise Psychology, 32*, s218–s219.
- Tenenbaum, G., Kamata, A., & Hayashi, K. (2007). Measurement in sport and exercise psychology: A new outlook on selected issues of reliability and validity. In G. Tenenbaum & R.C. Eklund (Eds.), *Handbook of Sport Psychology* (3rd ed., pp. 757–773). Hoboken, NJ: John Wiley & Sons.
- Thomas, P.R., Murphy, S.M., & Hardy, L. (1999). Test of performance strategies: Development and preliminary validation of a comprehensive measure of athletes' psychological skills. *Journal of Sports Sciences, 17*, 697–711.
- Vealey, R.S. (1988). Future directions in psychological skills training. *The Sport Psychologist, 2*, 318–336.
- Vealey, R.S., & Garner-Holman, M. (1998). Applied sport psychology: Measurement issues. In J.L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 433–446). Morgantown, WV: Fitness Information Technology.

- Willis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. London, UK: Sage.
- Woodcock, C., Holland, M.J.G., Sharp, L., Duda, J.L., & Cumming, J. (2010a). Psychological skill use in adolescents: Exploring the structural and temporal validity of the TOPS. *Journal of Sport & Exercise Psychology*, 32, s233.
- Woodcock, C., Holland, M.J.G., Sharp, L., Duda, J.L., & Cumming, J. (2010b, October). Origins and evolution: Coming full circle with mental skills training. In J. L. Duda (Discussant), *An action research approach to mental skills training: Experience, evaluation, and evolution*. Symposium conducted at the meeting of the Association for Applied Sport Psychology, Providence, RI.
- Young, B.W., & Starkes, J.L. (2006). Measuring outcomes of swimmers' non-regulation during practice: Relationships between self-report, coaches' judgments, and video-observation. *International Journal of Sports Science & Coaching*, 1, 131–148.