JOURNAL OF APPLIED SPORT PSYCHOLOGY, 17: 1–19, 2005

ISSN: 1041-3200 print / 1533-1571 online
DOI: 10.1080/10413200590907531

Taylor & Francis
Taylor & Francis Group

Tracing the Development of Athletes Using Retrospective Interview Methods: A Proposed Interview and Validation Procedure for Reported Information

JEAN CÔTÉ

Queen's University

K. ANDERS ERICSSON

Florida State University

MADELYN P. LAW

Queen's University

A new interview procedure is proposed for collecting valid information on the acquisition of high-level performance in sport. The procedure elicits verifiable information on the development of athletes' achievements in their primary sport, as well as factors that might influence performance, including involvement in other sporting activities, injuries, physical growth and quality of training resources. Interviewed athletes also describe their engagement in specific training and other relevant activities during each year of their development as well as how they experienced each type of activity. The collected information is then examined to identify those aspects of the athletes' recall of their development that meet criteria of reliability and validity. Recommendations to coaches and scientists are discussed for how retrospective interviews can uncover aspects of development that distinguish elite from less accomplished athletes.

Received 27 December 2003; accepted 28 August 2003.

The order of the first two authors of the manuscript is arbitrary because both contributed equally and extensively to the preparation of this manuscript. Support for the writing and data collection for this manuscript was given by standard research grants from the Social Sciences and Humanities Research Council of Canada (SSHRC Grant #410-99-0525 and Grant #410-2002-0325) to the first author and by funds from the Conradi Eminent Scholar Endowment of the Florida State University Foundation given to the second author. Sections of this manuscript were written when Jean Côté was a visiting academic (March 1–June 30, 2002) in the department of Human Movement Studies at The University of Queensland, Brisbane, Australia. Financial support to work on this manuscript as a fellow at the Center for Advanced Study in the Behavioral Sciences during 2002–2003 was provided by the John D. and Catherine T. MacArthur Foundation, Grant #32005-0 to the second author. The authors extend their gratitude to Whitney Sedgwick for her contribution to the development of the interview guide. The authors also gratefully acknowledged the constructive comments on earlier drafts of this manuscript by Bruce Abernethy, Ray Amirault, Joseph Baker, Kristin Côté, David Eccles, Len Hill, Jennifer Robertson-Wilson, John Salmela, and Carla Urena.

Address correspondence to Jean Côté, Ph.D., Queen's University, School of Physical and Health Education, Kingston, Ontario, Canada, K7L 3N6. E-mail: jc46@post.queensu.ca

In their pioneering research, Bloom and colleagues (1985) interviewed tennis players and swimmers who had exhibited elite performance by winning international competitions, such as Wimbledon and the Olympic games. Based on interviews with these athletes, their parents, and coaches, Bloom and colleagues (1985) inferred a general pattern of development that appeared necessary to reach this ultimate level of elite performance. For instance, the elite athletes reported beginning their sport participation at an early age and throughout the course of their development they had access to high quality coaches and superior training environments. Bloom and his colleagues argued that the developmental conditions and the level of engagement in domain-related activities differed dramatically for elite athletes compared to average children. However, Bloom (1985) did not systematically study the factors that could account for individual differences in attained performance among top-level athletes with similar developmental opportunities.

In a subsequent study investigating another performance domain, music, Ericsson, Krampe and Tesch-Römer (1993) analyzed the current level of training and the developmental histories of several groups of highly accomplished musicians to account for individual differences in their current level of achievement. These musicians identified one type of practice activity, *practice alone,* which was related to the attained level of performance of expert musicians. When assigned attainable goals for improving specific aspects of their performance, the students would engage with full concentration in associated problem solving and practice alone activities—this type of designed practice was called *deliberate practice* (Ericsson et al., 1993). The music students who engaged in solitary deliberate practice were thus able to control the structure, detail, and duration (daily and weekly) of the practice. Ericsson et al. (1993) showed that the reported duration of this form of effective deliberate practice was related to the level of attained music performance. Based on reviews of expert performance in many domains including sport, Ericsson and his colleagues (Ericsson, 1996; Ericsson et al., 1993; Ericsson & Lehmann, 1996) argued that expert performance and associated characteristics (e.g., superior tapping speed in expert typists, greater flexibility of fingers in expert musicians, and larger hearts in endurance athletes), should be viewed as *consequences of attaining a sequence of increasingly challenging goals through extended deliberate practice*, rather than the presence of any innate talent.

Recent studies have asked athletes at different levels of achievement to estimate how much time they have engaged in practice related activities during their development. Hodges and Starkes (1996) distributed a mail-in questionnaire to a group of international wrestlers and a group of club-level wrestlers where the athletes were asked "to think back to the amount of time they had spent practicing for wrestling: alone, with others, in practice-related activities and in everyday activities during a typical week" (pp. 406–407). They found that international level wrestlers reported increasing their amount of practice time more rapidly than club wrestlers. After wrestling for 6 years, the future international wrestlers were engaging in reliably more weekly practice than wrestlers that only reached the club level. A more detailed analysis, where practice alone was distinguished from practice with others, showed that only practice with other athletes was reliably higher for the international wrestlers when compared to club wrestlers. In order to identify the specific activities that would have the characteristics of practice alone in music, the wrestlers were also asked to rate different aspects of a large number of specific practice and everyday activities (similar to those rated by the musicians interviewed by Ericsson et al., 1993). The activities were rated for "relevance to improving wrestling performance" and "how enjoyable the actual activity was" (Hodges & Starkes, 1996, p. 407). The wrestlers were also asked to rate the activities with regard to the required physical effort and mental concentration—an improvement from the Ericsson et al. (1993) study where the musicians only gave a single rating of general effort. The ratings did not differ between the two groups of

wrestlers. Wrestlers generally rated their level of concentration in practice activities, relevance to improvement, and enjoyment of some of the activities as "high." These patterns of rating characteristics of practice activities have been found with figure skaters (Starkes, Deakin, Allard, Hodges, & Hayes, 1996) martial artists (Hodge & Deakin, 1998), as well as soccer and field hockey players (Helsen, Starkes, & Hodges, 1998).

One of the most generalizable findings about practice across different domains of expertise is that its weekly duration increases gradually during the development of expert performance (Starkes et al., 1996). During the first couple of years after the introduction to the domain, future expert performers report spending, on the average, just over five hours per week in practice—regardless of their age of introduction to the domain. A decade later, the overall duration of practice has increased to almost 25–30 hours per week (Starkes, 2000). Recent studies (Baker, Côté, & Abernethy, 2003a; Côté, 1999; Soberlak & Côté, 2003) have shown that not only does the amount of practice increase but also the structure and type of activities change as a function of the age of the child athlete. Accordingly, Côté and colleagues (Côté, 1999; Côté, Baker, & Abernethy, 2003; Côté & Hay, 2002) have proposed the Developmental Model of Sport Participation as a way of describing the developmental activities of athletes and account for individual differences in their current level of participation and performance.

The interview procedure proposed in this article was developed to trace the development of expert performance in sports (Côté et al., 2003) and to allow assessment of deliberate practice within the expert-performance framework (Ericsson & Smith, 1991). At this time, various procedures ranging from open-ended interviews (e.g., Bloom, 1985) to fixed-response questionnaires (e.g., Starkes et al., 1996) have been used to examine the activities that promote sport expertise throughout the lifespan. However, no standardized method with high reliability, high validity, and high discriminatory power has been proposed to collect data about the development of expert performance in sport. In presenting this method, we first describe how to collect verifiable objective measures of performance and measure improvements in performance at the most detailed level possible. Second, we propose methods for eliciting information about the amount and structure of practice-related activities and training resources as a function of development of the athletes. An interview procedure is proposed to map out longitudinal changes of sport participation patterns and various training activities. This detailed account of longitudinal changes will improve our understanding of how different activities contribute to the development of expertise at various stages of athletes' involvement in sports. Third, we describe how our interview procedure is designed to allow us to assess the reliability and validity of all information reported by the athlete. Hence, in a separate section we will review findings about which type of retrospective information the athletes can accurately recall many years or even decades later.

The development of athletes at different levels of achievement is sufficiently diverse to make it very difficult to capture relevant information via standardized questionnaire. Therefore, the interview procedure described in this paper allows the researcher to adjust the questions to the particular background and development of the athletes, and yet collect information in a standardized manner. The development of expert peformance in sport usually starts in childhood, yet our understanding of factors such as stages of development, training, health status, and resources to achieve long-term excellence rather than short-term burn out are limited. Throughout their development, athletes progressively develop physically and mentally to meet the increasing demands of competition and performance. Describing the process of development in sport from early childhood to adulthood will help develop informed practices and policies for sport programs. The careful analysis of the lives of athletes of different level of performance will uncover valuable information about the optimal conditions for learning

and will provide athletes, parents and coaches with guidelines on how to maximize learning and participation at various stages of an individual's involvement in sport.

DETAILED RETROSPECTIVE INTERVIEW PROCEDURE

With this interview procedure, we are interested in retrospectively assessing how elite athletes' development may have differed from that of less accomplished athletes in the same sports and events. Our interview procedure[1] was designed under the assumption that when individuals answer interview questions based on recall of past episodic experiences, these individuals will be more accurate and reliable than when they are forced to infer and reconstruct answers to general questions (see Ericsson & Simon, 1993, for the theoretical rationale). Furthermore, answers to specific questions about events and experiences can often be evaluated for accuracy, whereas the validity of general answers and evaluations is difficult, if not impossible, to assess with objective methods. For example, the questionnaire introduced by Starkes et al. (1996) and used in several expertise studies asked athletes to rate the relevance that specific training activities had on their development. Asking athletes, especially young athletes, to make this kind of inferential evaluation is problematic since it would be difficult for anyone to validly assess which of the many training activities were responsible for the improvements in their performance at a given age. Therefore, these judgments can only reflect the athletes' current beliefs or opinions about their development. It is also important to note that our interview procedure does not ask athletes to remember the "real time sensorimotor demands" (Beilock & Carr, 2001, p. 704) of their performances but rather focuses on the recall of factual knowledge about concrete activities they engaged in throughout their development. Our interview procedure focuses on gathering longitudinally indices of performance as well as the types and amount of activities that participants engaged in throughout their development. Furthermore, the various conditions associated with each activity are assessed.

The first section of the interview was designed to assess the specific level of performance the athlete achieved in their sport ranging from regional to the international level. This information provides a detailed outline of the ages at which the athlete attained higher levels of performance. The general goal is to obtain a developmental profile of how athletes' performance changes throughout their development. This section of the interview yields interesting information as to the ages when major events occurred in the athletes' development and helps identify objective performance criteria that may differ between groups of athletes of various expertise levels. This section of the interview usually takes between 30 minutes and one hour and could be conducted independently from the other two sections.

Sections two and three of the interview follow a similar format. The information gathered in these two sections used a series of charts with variables arranged in a row across the top and columns referring to chronological age and activities engaged in. The interviewer used specific wording to elicit required information in a standardized way. Similar to a method recently suggested by Massey (2000), each chart is organized around activities that the athlete was involved in, giving coherence and order to the interview. For each chart, the respondent began at an appropriate point in their life and moved chronologically forward in time by quantifying the impact of their involvement in different activities. When information about one variable had been gathered by reaching the present, the next variable was considered in a parallel fashion**.** These two sections of the interview last between two and

[1] A copy of the interview procedure is available upon request to the first or second author.

Table 1
Content of Interview

1. Measures and Description of Current and Past Levels of Performance
   For each year of the athlete's engagement in her/his main sport the following variables are elicited:
   - Individual and/or team performance (e.g. running time)
   - Sport specific milestones (e.g. win-loss record, selected for an all-star team, etc.)
2. Engagement in Domain-Related Activities
   For each year of the athlete's engagement in her/his main sport the following variables are elicited:
   - All physical or mental training activities related to main sport
   - Number of hours spent in each activity per week
   - Number of months per year
   - Enjoyment of each activity
   - Physical effort
   - Mental concentration
3. Factors Limiting the Quality and Quantity of Training
   a) Involvement in sporting activities other than primary sport
      For each year of the athlete's engagement in sport the following variables are elicited:
      - All organized or unorganized sports or physical activities
      - Number of hours spent in each activity per week
      - Number of months per year
   b) Height and weight
      Athlete recalls his/her height and weight, and especially changes from year to year.
   c) Quality of training resources
      For each year of involvement in his/her main sport the athlete is asked to provide a composite rating of the quality of the training resources available.
   d) Health/injury
      For each year of involvement in his/her main sport the athlete is asked to describe the nature and duration of any injuries and to provide a rating of his/her overall health.

three hours and provide a longitudinal and detailed account of a participant's involvement in sports.

The three general content areas of the interview procedure are outlined in Table 1 along with the variables assessed in each content area. A longitudinal analysis of the three categories of variables outlined in Table 1 shed light on two fundamental questions about sport development: 1) How does performance change over time? and 2) Can we predict differences in performance? The first question is descriptive and focuses on each athlete's pattern of performance changes over time. The second question is relational and is aimed at examining variables that can explain changes in performance within and between athletes. By proposing a retrospective approach that acknowledges both changes in performance over time and factors that can affect these changes, it will be possible to better articulate the nature of stages of development in sport.

Content Area One: Measures and Descriptions of Current and Past Levels of Performance

To study the development of expertise, investigators typically search for domains in sport where longitudinal measurements of the same individuals' performance on the same events are available. In order to study stable improvements in performance, it is necessary to compare the performance across competitive seasons. For example, in a study with international-level rowers (Sedgwick, Ericsson, Beamer, & Côté, 1999), one rower reported that at age 25 he was participating at the elite international level, had won a silver medal at the World Championships, and consistently completed 2000-meter races in 6.15 minutes. An account of his performance

on the 2000-meter race showed that his time decreased gradually from 6.44 minutes at age 17 to 6.15 minutes at age 25. As can be seen, any attempt to explain improvement becomes much easier when the associated changes in measured performance on the same event are given longitudinally.

In certain sports (especially team sports), it is sometimes difficult to measure athletes' performances throughout the course of skill development. Although elite-level teams utilize coaches and sport scientists hired specifically to collect individual player's performance statistics, most youth sport teams do not keep records of individual athletes' performances. When objective, longitudinally-based performance measures of individual athletes are not available, the individual performance at a given point of development can only be inferred by an athletes' relative standing within a team (first-string versus second-string players), the general level of performance of the team and/or a coach evaluation of the athlete. It is also possible to obtain verifiable information about athletes' transfer to other teams. When players transfer successfully to teams playing at a higher level of competition (such as from local to regional, from regional to national, or from national to international), then we should be able to infer that these players met the demands required of the higher level of competition. Therefore, in our interview, each athlete is asked for the age when they first participated in organized competitions at the club, district, provincial, national and international levels. In addition, athletes are asked about general developmental events, such as the age when they made the decision to attempt to become elite-level athletes.

Content Area Two: Engagement in Domain-Related Activities

The second step of the interview procedure consists of collecting information about the type and level of the athletes' engagement in specific training activities. First, the athletes are asked to list all physical or mental training activities related to the main sport in which they engaged during their development. To facilitate this process, the interviewer asks probing questions designed to improve the athletes' recall of training activities and sport-related activities. For example, participants are asked to list activities from various training categories such as organized training activities (e.g., training with a team), self-initiated training activities (e.g., weight training), and individualized instruction (e.g., private lesson with a coach). Once the list of activities has been completed, each athlete is asked to estimate the quantity and the quality of involvement in each activity during each age in a manner described below.

Estimation of the Number of Hours Per Year

For each year of involvement in the sport, athletes are asked how many months of the year they were engaged in each training activity. For the portion of the year that the athletes were active, athletes are asked to estimate the number of hours that they engaged in the activity in an average week. Based on these two sources of information, the interviewer can calculate the number of estimated hours per year that the athlete has been involved in each activity.

Enjoyment of Engaging in Activities or the Experiences of Fun

According to Ericsson et al. (1993) engagement in deliberate practice requires effort, generates no immediate rewards, and is motivated by the goal of improving performance rather than inherent enjoyment. In a review of the literature on play and practice activity in sport, Côté et al. (2003) proposed a series of dimensions, including inherent enjoyment, that helps discriminate between play and practice activities. However, the measurement of the "inherent enjoyment" of an activity has followed various procedures and consequently has produced mixed results. Ericsson et al. (1993) instructed their expert musicians to ignore the consequences of an

activity and to focus on the inherent enjoyment of the activity itself. On the other hand, the questionnaires used in the sport research typically have instructed participants to evaluate "the enjoyment derived from the actual activity" (Helsen et al., 1998; p. 18). The latter instructions might measure a different construct, because participants might confound the enjoyment of the consequences of participating in the activity (improved performance) with the concurrent enjoyment of the activity itself (Ericsson, 1996). However, neither of these types of instructions would allow an external observer to rate how much a given individual enjoys engaging in the different training activities. Consequently, we asked athletes to rate how much fun (or enjoyment) they experienced in each activity. We hoped that this emotion would be expressed in the athletes' faces and posture (Bartlett, Hager, Ekman, & Sejnowski, 1999), and thus observable for parents and coaches.

To increase the control over how the interviewed athletes generated the ratings of fun, they were asked to think back to a given age and recall the activity with the most fun, such as watching a movie, playing video games, or celebrating a birthday. Thus, for each age the athletes recalled and described the activity with the most fun, and were asked to consider that activity to be 100% fun. Using this activity as a reference, they were then asked to rate how much fun it was to participate in each training activity during that year of development. For example, data from an interview with a rower showed ratings of fun at age 18 and 19 at 90% for on-the-water training, 70% for weight training, and 40% for calisthenics. The rower rated going to a party with friends as 100% fun when he was 18 and 19 years old, which was then used as his reference point at this age to rate the different training activities of rowing.

Physical Effort and Mental Concentration

In many sports, the effort involved in sustaining mental concentration while executing performance and engaging in tasks necessary to improve aspects of skilled performance are not the only factors impacting performance (e.g., Ericsson et al., 1993): *physical effort* to execute motor performance during competitive events (and training tasks designed to attain and maintain physiological adaptations) also plays a significant role (Starkes et al., 1996). Following a procedure similar to that for ratings of fun, athletes generated ratings of physical effort of their training activities for each year. They first identified the most effortful activity and the specific year when they had exerted the highest level of physical effort on a consistent level. The athletes were asked to consider this level of effort to correspond to 100% physical effort. In addition, they were asked to identify an activity where the physical effort level had been non-existent or at its lowest level and they were asked to consider this level equal to 0% physical effort. Using these two points of reference, the athletes rated their physical effort level for all training activities for each year of involvement. As an illustrative example, one of the interviewed rowers reported that he experienced the maximal physical effort (100%) for on-the-water training, ergometer training, and ergometer testing at age 18 and 19. On the other hand, watching television was rated as 0% physical effort at age 18 and 19. Using these reference points this rower rated the physical effort for all the other activities. For example, at age 20 he rated his physical effort for on-the-water training to be 65%. His rating of physical effort for ergometer testing was 75%. To assess mental concentration each athlete identified an activity associated with maximal (100%) and minimal (0%) mental concentration and rated each training activity for each year of involvement using the associated scale.

For the ratings of physical effort, mental concentration, and enjoyment, each athlete used a single reference (i.e., their most intense year of training or most enjoyable activity). This rating procedure allows us to infer not just which training activity was rated as more effortful (or enjoyable) in a given year but also the age when the athletes' rated effort (or enjoyment) associated with a training activity increased. Our rating procedure has important limitations

in that it does not allow us to compare directly ratings between different athletes, as each athlete may have used different experiences and intensity levels at a reference point of 100%. The use of a ratio scale from 0% to 100% may appear to be overly sensitive to measure subjective states such as physical effort, mental concentration, and enjoyment. However, this procedure allows participants to consider their responses relative to the complete absence of each variable and to rank order their involvement in various activities considering the absolute magnitude of each variable. It is also possible, at least in principle, to search for everyday activities, such as watching TV, cleaning one's room, or reading a difficult textbook, where the experienced level of concentration and physical effort is likely to be very similar across all participants. By finding the appropriate transformation of the observed ratings, it is possible to match values on these standard activities and then compare the transformed rating scores across participants.

Content Area Three: Factors Limiting the Quality and Quantity of Training

Singer and Janelle (1999) identified five general factors that may constrain athletes' abilities to reach their full potential that is: Personality traits, intelligence, information processing processes, physical characteristics, and situational circumstances. The methodological challenge becomes to have a sensible metric that would allow the measurement of these variables retrospectively with validity and precision. Singer and Willett (2003) recently suggested that to be analyzed longitudinally a variable must be "equatable" over time—that is "a given value of the outcome on any occasion must represent the same "amount" of the outcome on every occasion" (p. 13). Because people do not usually keep records over time of their personality traits, intelligence, or information processing processes, it becomes almost impossible to assess these retrospectively. Similarly, there are limits to the "physical characteristics" and " situational circumstances" associated with one development in sport that can be recalled reliably and validly. Therefore, our procedure assessed the following "physical characteristics" and "situational circumstances" that can possibly limit the quality and quantity of training: 1) involvement in sporting activities other than the primary sport, 2) height and weight, 3) quality of training resources, and 4) health/injury.

Involvement in Sporting Activities Other than Primary Sport

Recent studies have shown that participation in competition, other relevant activities, and "deliberate play" activities benefit the development of expertise in certain sports (Baker et al., 2003a; Baker, Côté, & Abernethy, 2003b; Soberlak & Côté, 2003). By collecting information about engagement in other sports activities, we will be able to address theoretical issues concerning transfer and specificity of training for the development of expert performance in sport. Following a procedure similar to that used in our interview to study domain-related activities in the primary sport, the athletes were asked to list all involvement in any organized or unorganized sports or physical activities throughout their development. Once a comprehensive list of sporting activities had been completed, the participants reported when they started and stopped participating in each activity, and estimated the number of hours per year of engagement in each activity.

Height and Weight

In many sports, increase in body size is associated with increase in performance. Given that children's and adolescents' bodies grow at different rates, it is important to monitor their growth from year to year. An analysis of the increases in body size will allow us to control statistically for the effects of increased body size on observed performance to get a more accurate estimate

of effects of training and experience. Therefore, athletes recalled their height and weight, and especially changes in height from year to year. If athletes were unable to recall their exact weight or height, they were asked if they were heavier/lighter or taller/shorter in comparison to other athletes of the same age in their sport.

Quality of Training Resources

The availability of master coaches and superior training resources is likely to facilitate the quality of development of performance for highly motivated individuals (Côté, 1999; Salmela, 1995). Our interview assessed the quality of an athlete's available training resources as a function of their development. We defined quality of resources as a composite factor that includes the amount of money invested on instruction and training, quality of the training facilities and equipment, and the quality of coaching and social support.

Given that the best possible training environments may differ for different stages of development, the athletes are also asked to identify the best training environment in the world for each age, and to use that environment as a reference point (equating that environment as 100%) for their respective ratings. A rating of 0% was defined as complete lack of resources. Athletes rated the training resources on a scale between 0% and 100% using these two reference points, regardless of whether the athlete had personal experience of the best training environments. For example, when one of the interviewed rowers was asked to think of the best possible quality of training resources that he or another rower could experience at age 18, he described a specific training environment where he would be able to row all year with access to superior quality of coaching. He then provided a rating of his own environment at age 18 as being 85% of this best possible environment. Athletes in a given sport generally know the locations of the best training centers for the development of athletes and are likely to use similar reference points for their ratings of training resources.

Health/Injury

Injuries are likely to influence athletes' performances in competitions. Any attempt to estimate increases in performance related to practice by comparing the observed performance across years must take into account the athletes' injuries for each year. Therefore, for every year of their involvement, the athletes are asked if they had ever sustained an injury that had an adverse effect on their performance in their main sport. If so, they are asked to describe the nature and duration of this injury. Probe questions such as "how did this injury occur?" are used to help the athlete recall the exact type and nature of the injury. Keeping the description of their injuries in mind, the athletes rate their health for each year of their development. The athletes are asked to rate their "health" on a scale, where a rating of 100% corresponds to a year when an athlete had not been bothered by injuries, while a rating of 0% corresponds to being unable to train and compete for that year. For example, one of the interviewed rowers reported that he sustained an injury that did not permit him to train at a normal level for three months in 1992. Due to the nature of the injury and the reduced level of training, the rower rated his health that year as being at 85%.

THE RELIABILITY AND VALIDITY OF RETROSPECTIVELY AND CONCURRENTLY RECALLED INFORMATION

It is obvious that the usefulness of our proposed methodology is critically dependent on the validity of the reported information. When the reports require recall of activities and events that took place months, years, or even decades ago, the accuracy of the reported information cannot be taken for granted, regardless of participants' motivation to provide accurate reports.

Research on memory has shown that longer retention intervals lead to lower accuracy of recall for virtually all types of memory at uniform and predictable rates, unless the information is accessed and rehearsed during the delay (Bahrick, Hall, Goggin, Bahrick, & Berger, 1994; Rubin & Wenzel, 1996). For example, students' memory of material mastered on final exams in introductory college courses is virtually lost after a 2–4 year delay. Only when the mastery of the material is cumulative and students take courses where each new course builds on the previously acquired knowledge, such as in course sequences in mathematics and Spanish (Bahrick, 1984; Bahrick et al., 1994), has substantial memory been demonstrated years and decades later. More generally, bodies of knowledge that have been thoroughly mastered, such as memory of friends and teachers in high-school (Bahrick, Bahrick, & Wittlinger, 1975) and street names from their childhood neighborhood (Schmidt, Peeck, Paas, & van Breukelen, 2000), are retained well into adulthood.

For recall of most other types of information, forgetting is substantial over extended time and accuracy of recall is relatively unreliable. It is even of greater concern to researchers that recall of information from past experiences is often systematically biased. For example, Bahrick, Hall and Berger (1996) found that recall of high-school grades was influenced by the desirability of the grade received—the grade of A was recalled accurately 89% of the time but the barely passing grade of D was only correctly recalled 29% of the time. In a review article, Ross (1989) showed that the reported memory in many studies is the result of reconstruction and inferences. Participants rely on their current feelings, attitudes and situations to extrapolate what they think they might have thought or experienced at earlier times. For example, Markus (1986) found that when participants were asked to recall their attitudes held at the time of a previously administered attitude test about issues, such as equality of women and legalization of marihuana, their recall of the previously held attitudes was better predicted by their *current* attitudes than by their original attitude responses recorded for that test. It is thus possible for respondents to be very reliable in their responses without the responses reflecting an accurate memory of their past experiences. If participants had similar beliefs about the typical development across the life span, then their judgments about inferred experiences as a function of age might show a very similar pattern without the need for mediation of accurate memory of recalled feelings and experiences (Ross, 1989). Another concern is that respondents can only recall a small number of vivid experiences (and which may or may not be representative), which may lead to biased generalizations. Research on vivid personal experiences, generally referred to as flashbulb memories (Brown & Kulik, 1977), has shown that the level of accuracy of these memories is higher than that for more mundane events. However, even vivid memories show normal forgetting with long delays and, most importantly, this research has shown that the experienced vividness of memories at the time of recall is not a valid indicator of high accuracy (Conway, 1995).

In light of the reviewed evidence on substantial forgetting and bias in long-term memory of activities and experiences, the retrospectively recalled and reconstructed information cannot simply assumed to be valid. In the following sections, we will discuss approaches to assess the validity and reliability of the information elicited in our interview procedure and review other studies that have collected similar types of information. For each category of information elicited in our interview procedure we will assess its reliability and validity.

Current and Past Levels of Performance

Information of an individual's performance in public competitions can nearly always be checked against public records. For example, information from the Canadian rowing association website provided indicators of performance that validated information obtained from an expert Canadian rower (Sedgwick et al., 1999). The website information allowed a comparison of

various indicators of performance for the rower for 13 years of competition from 1986 to 1998. The information reported by the rower was consistent with the information obtained on the website for the 13 years assessed.

In cases where the individuals' performance is not permanently recorded by organizations, there are other independent methods of validation. Present and past coaches as well as parents should be able to validate athletes' achievement at various stages of their career. When the athlete's interview is conducted independently of the interviews with the coach or parent, investigators can compare the recalled information to assess convergent validity (Baker et al., 2003a; Soberlak & Côté, 2003).

Height and Weight

The height and weight of athletes are often part of their personal profiles and are readily available from websites of teams or from documents available from the various sporting federations. For example, in a study of elite rowers (Sedgwick et al., 1999) the retrospective estimate of height and weight reported by one of the rowers was confirmed by a personal profile available from the national rowing association. In some sports there are implications of changes in weight for the class of competition in sports, such as wrestling, boxing, and weight lifting, and physical growth is likely to lead to changes in equipment, such as clubs and skates. More generally, height and weight are sufficiently important to the social interactions among teenagers that it would be reasonable to assume that large changes in height during a growth spurt or sudden increases in muscle mass would be distinctive and easy to remember.

Quality of Training Resources and Injuries

Athletes' health and their training environments refer to relatively observable and publicly available states of affairs. Hence, we would expect that parents' and coaches' ratings of these variables would correspond closely to the athletes' ratings of these two variables as a function of the athletes' age. Our interview procedure has recently been applied in a study of 15 expert Australian team sport athletes (Abernethy, Côté, & Baker, 2002). Thirteen parents of the fifteen athletes were asked independently to assess the quality of training resources of their child for each age throughout their development in sport. The results showed a high level of consistency between the parents and the athletes' assessment of quality of resources throughout development.

Major injuries and other major changes in health that influence performance and ability to practice should be memorable for athletes. For example, Beamer, Côté and Ericsson (1999) found that recall for the health/injury status of a gymnast with major injuries was very consistent when measured two years apart. Whenever possible one should, however, strive to find objective data to validate the recalled information, such as medical records or restrictions on competition and performance. More generally, structured interviews of large samples of middle-age adults about their lifetime exercise have shown relatively low (around $r = 0.4$) reliability of reports of exercise-related injuries when individuals were interviewed four to five years later (Ropponen, Levalahti, Simonen, Videman, & Battié, 2001). However, these reliabilities refer to reported memory for events from as much as twenty to thirty years ago by mostly non-athletes and therefore should be viewed as a lower bound for validity of reports on injuries by young athletes.

Concurrent and Past Engagement in Sporting Activities and Domain-Related Activities

The early work on the development of music performance focused on concurrent assessment of domain related activities, such as practice alone, public performance, and playing music

with others for fun. Given that no single activity in sports captures deliberate practice as completely as practice alone in music, investigators have tried to assess the weekly duration of many different activities to identify the type of activity that best captures the characteristics of deliberate practice (Starkes et al., 1996). The most convenient method to assess the weekly duration of current and past activities involves instructing the athletes to fill out questionnaires. However, it is necessary to assess the reliability as well as various forms of converging validity for the reported information. Given that reliability is a pre-requisite for validity, we will review some evidence for test-retest reliability before reviewing evidence on convergent validity with estimates by parents and with athletes' activity diaries.

Test-Retest Reliability of Past Practice

Helsen et al. (1998) administered the same questionnaire, six months apart, for assessing retrospective estimates of weekly duration of individual and team practice activities to a sample of 10 soccer players. The test-retest reliability was uniformly high and reached statistical significance in spite of the small sample. Similarly, Hume, Hopkins, Robinson, Robinson and Hollings (1994) found high reliabilities for estimated amounts of cumulative and current training (average $r = 0.84$) when they re-tested a sub-sample of eighteen rhythmic gymnasts within 10 weeks of the original administration of the questionnaires. Furthermore, research on the lifetime exercise history of larger representative samples of middle-aged adults ($n = 150$) has shown that the mean hours of exercise per week had the highest reliability (intra-class correlation coefficients around 0.8) of all the information elicited at the re-test several years later (Ropponen et al., 2001). Another large sample study ($n = 115$) has found correlations in the 0.7 range for test-retest reliability of reported lifetime exercise activity (Friedenreich, Courneya, & Bryant, 1998).

Convergent Validity of Estimated Concurrent Level of Domain-Related Activity and Diaries

Research on recent physical activity (Weston, Petosa, & Pate, 1997) has shown that high-school students ($n = 90$) were very consistent in recalling their physical activity during the previous day ($r = 0.98$) when asked to recall the same information an hour later. More importantly, the recalled information showed a high agreement (around $r = 0.8$) with concurrent data on physical activity collected by pedometers worn by a subset ($n = 48$) of the students. When duration of activities that are recalled daily in diaries are compared to estimates of weekly duration, it is essential that the estimates refer to the same activity and time period.

More generally, the level of daily activities and practice may vary substantially across time as a function of the athletes' development. Depending on injuries and one's competitive involvement, athletes' level of weekly practice may differ substantially. Hence, it is important to differentiate the poor reliability of estimates of duration of practice for the same time period from variability of the amount of weekly practice during different periods of the athletes' development. This distinction cannot be made in some studies where the level of estimated practice for a recent diary week is correlated with the total amount of practice during the athletes' entire development. For example, Helsen et al. (1998) found high correlations between diary estimates of practice for a current week with the total amount of estimated practice accumulated during development. In contrast, Hodges and Starkes (1996) failed to find reliable correlations between diaries and estimated total practice during one's development. In sum, only when the estimates of weekly duration of activities refer to the same time period will the calculated correlations reflect the reliability of the performers' ability to accurately estimate the duration of the rated activities.

Convergent Validity of Estimates of Engagement in Practice by Athletes and Their Parents

Coaches, parents, or training partners usually have knowledge of the quantity of training that athletes have invested in their sport. Interviews with them can therefore be used to assess the validity of the information collected from the athletes. For example, thirteen parents of the fifteen Australian team sport athletes were asked independently to estimate whenever they had enough knowledge of their child's training routine, the number of training hours that their child had spent for each year of the child's development. After the age of 19, many parents felt they could not provide accurate information due to the fact that their child left home for school or a national training center. In spite of the range restriction of the individual differences in practice for this elite sample of athletes, the correlations between athletes' and parents' assessments of training hours were statistically reliable for all ages between 12 and 19.

In sum, the reviewed evidence suggests that elite performers and athletes are able to reliably estimate the number of hours they spent in sporting and training activities at different periods in their development. Perhaps the most compelling reviewed evidence for reliability/validity comes from the agreement of parents' and athlete's estimates of training levels. However, the most promising method of evaluation will be parents and athletes who have kept training logs for long periods of development. These training logs could be especially interesting with retired athletes who have not recently reviewed them; the ability of these athletes to recall their training levels can then be checked against these logs.

Enjoyment, Effort, and Concentration

Researchers in sports have attempted to distinguish deliberate practice from other types of practice activities by asking athletes to rate their enjoyment, physical effort, and mental concentration during various types of domain-related and everyday activities (Helsen et al., 1998; Hodge & Deakin, 1998; Hodges & Starkes, 1996; Starkes et al., 1996). None of these studies have assessed the validity of the athletes' ratings of fun, effort, and concentration. Based on the results of a case study, we have reason to believe that even the reliability of these ratings may be questionable. In this case study, we asked a gymnast who had rated various activities as part of Beamer et al.'s (1999) study to return after two years and give the same ratings again. An examination of the ratings showed no reliable consistency across the two test occasions, especially for the ratings of fun, but also for the ratings of effort and concentration. In addition, Ropponen et al. (2001) found that subjective ratings of exercise intensity had the lowest re-test reliability of all the parameters collected from lifetime exercisers. Although no firm conclusions should be drawn from this limited evidence, it raises the possibility that these ratings may not meet minimum standards of reliability.

A careful examination of the task of rating the enjoyment, effort and concentration of different activities reveals some challenges for validation. In studies on deliberate practice in sport (Helsen et al., 1998; Hodge & Deakin, 1998; Hodges & Starkes, 1996; Starkes et al., 1996), athletes were instructed to give an overall rating of enjoyment, effort, and concentration for different activities. This kind of rating would reflect athletes' experience of the activity in general by implicitly aggregating many years of accumulated experience of that activity. Given that we cannot know whether the difficulty of generating reliable ratings is due to poor memory of episodes or variability of the process of aggregation of many experiences, it would make sense to start by assessing the reliability of reported memory for individual experiences of a given training activity. The challenge in assessing more subjective and less observable variables such as fun, effort, and concentration involves finding observable indicators that can be used as valid measures of these psychological states in specific episodic events. However,

it is difficult to find objective indicators of athletes' subjective levels of effort, concentration, and enjoyment.

Objective Indicators of Physical Effort, Mental Concentration, and Enjoyment

The most widely accepted objective indicator of physical effort is elevated heart rate. Subjective ratings of physical effort have a remarkably close correlation with the individuals' heart rate even when highly trained athletes are studied in the laboratory using exercise bikes (Borg & Ottoson, 1986). This methodology can be extended to monitoring the effort level of training activities where the heart rate serves as an indicator of the physical effort associated with an activity (Noble & Robertson, 1996). It would be possible to assess the validity of self-reported measures of physical effort of specific training activities during a diary week by asking athletes to wear commercially available heart rate monitors during their training bouts (Hue, Le Gallais, Chollet, & Préfaut, 2000).

The most commonly used method to rate an athlete's level of mental concentration involves observation of the athlete's behavior. When athletes sit down to rest, joke around, or observe others train, one can infer a lower level of mental concentration than when these athletes are involved in practice activities. Micro-analyses of athletes' practice behavior (Deakin & Cobley, 2003; Starkes, 2000) shows that athletes spend a relatively small proportion of practice time on very challenging practice activities that lie at their current limits of performance and thus would require maximal levels of concentration. Objective descriptions of athletes' practice behavior can then be related to their ratings of mental concentration made immediately after a given practice session.

Concurrent observable indicators of enjoyment and fun have been more difficult to find and the best available methods may involve behavioral indicators such as facial expressions. Facial expressions of basic human emotions such as happiness and fun have been shown to reliably reflect subjectively reported emotion (Ekman, 1992; Smith & Crabbe, 2000). To obtain a measure of facial expressions, athletes could be videotaped in various training activities, and their ratings of fun, given immediately after the end of each training activity, could be compared against ratings made by others observing the videotapes, such as parents and coaches. If athletes rate their enjoyment of training activities several times during different training sessions it will be possible to examine the stability and reliability of their ratings of particular training activities. The averages of these immediate ratings can then be used to assess the validity of traditional ratings of enjoyment of training activities elicited by questionnaires and interview questions. All in all, it is important to corroborate self-reported measures of subjective experiences with behavioral or physiological indicators of the assessed characteristics to show that these ratings are reliable and valid when the problems of memory are minimized.

Rating Characteristics of Aggregated Experience over Extended Time

The practice of asking athletes to rate their overall experience of effort, concentration, and fun for a given activity aggregated across an extended time period raises several issues. With respect to accuracy of recall of past experiences, research on memory of emotions (see Kihlstrom, Eich, Sandbrand, & Tobias, 2000, for a review) shows that memory of intensity of emotions is especially vulnerable to forgetting, and particularly poor for unpleasant emotions (Thomas & Diener, 1990). The rapid forgetting of emotional experiences of recurrent activities, such as practice, would suggest that researchers should prioritize ratings of emotional experiences made immediately after the completion of an associated activity. However, this type of approach cannot be extended to retrospective interviews of athletes' development, where athletes report overall ratings of these aspects that integrate an athlete's experiences accumulated over hundreds or thousands of hours for a specific time period in their development.

Our approach attempts to uncover stable characteristics of the athletes' experiences of activities and, in particular, to describe stable changes in the athletes' perception of their physical effort, mental concentration, and fun over their development. Only stable long-term changes in these perceived characteristics are likely to lead to increased physiological adaptations and associated improvements in performance. For example, increased physical effort during an individual session of interval training or of lifting of weights will have an effect only when these increases in intensity are sustained in subsequent training sessions for the following months and years. Our assumption is that the successful attainment of increased levels of effort and concentration during training require deliberate and stable changes in training routine and subsequent recovery and rest by athletes (Ericsson, 2001). Consequently, when athletes recall experiences of a training activity from a given time period, they are likely to recall the experiences and circumstances associated with changes and the attainment of a new stable adaptation. For example, athletes may be able to recall at what age they first started to train under adverse weather conditions (important because training in such an environment indicates a high level of motivation to train). Ericsson et al. (1993; Krampe & Ericsson, 1996) used "taking naps to recuperate from practice" as an indicator of effort. Athletes' willingness to adapt their practice methods to accommodate constraints set by training injuries might be another indicator of their motivation to improve at the expense of immediate enjoyment. The latter type of behavioral evidence would be observable by parents, coaches, and training partners, and thus would allow validation by independent observers.

In our interview schedule, we have tried to develop questions and rating procedures that should induce recall of relevant specific episodic memories and details. Drawing on the "critical incident method" (Flanagan, 1954) and the methodology of collecting retrospective reports of activities as they were experienced (Ericsson & Simon, 1980, 1993), we encourage individuals to recall specific episodic events. Consequently, our interview schedule includes specific wording strategies to induce better recall (Bradburn, 2000; Massey, 2000; Menon & Yorkston, 2000). Our interview procedure uses probes to reconstruct specific events and the use of reference points that relate specifically to the respondent's life. To facilitate the process for generating ratings for effort and concentration, respondents are first asked to break up their career into stages using salient temporal boundaries. Respondents are then asked about activities that required the most effort (or concentration) at each of these stages, and to remember an event in which they exerted maximal effort. Once respondents have established activities that required the most effort during a certain period of their career, they are asked to compare the same activities between different periods (e.g., weight training at age 18 vs. weight training at age 22). This process demarcates a reference period and differentiates between activities.

In summary, psychological constructs such as fun, effort, and concentration are far less precisely defined and more difficult to measure than variables such as the number of hours spent training or the quality of training resources and injury. The reliable and valid measurement of these more subjective variables remains a challenge. Therefore, it will be important that future studies on expert performance in sport develop and refine methods for assessing the cognitive and emotional aspects that differentiate deliberate practice from other types of domain-related activities.

SUMMARY AND GENERAL CONCLUSION

As long as we are not able to predict accurately which young athlete will eventually reach the highest level, these outstanding athletes can only be distinguished after the fact. Consequently, retrospective interviews with such outstanding athletic performers will remain one of the primary sources of information on the acquisition of the highest levels of performance for

the foreseeable future. In this article, we proposed an approach designed to elicit information about the development of expert levels of performance that could be assessed with respect to accuracy, validity and re-test reliability. Only when we have developed procedures to elicit valid information about the development of elite performance will we be able to identify those aspects that distinguish elite from sub-elite athletes and derive the associated implications for selection and training.

The search for relevant information about athletes' development will be constrained by methodological factors, such as verifiability, reliability and validity, but also by plausible theoretical models. The design of our interview procedure was guided by the theoretical framework of deliberate practice (Ericsson et al., 1993; Ericsson, 1996, 2003) and the Developmental Model of Sport Participation (Côté, 1999; Côté et al., 2003; Côté & Hay, 2002). The primary focus is on the development of the athletes' performance in their primary sport, documented by their recall of their objective performance at different ages, along with reports of their success at different levels of competition at the district, regional, national or international levels. The athlete is then asked to recall the engagement in different types of training activities for each age and any information about other factors that could affect performance at given ages, such as access to superior training resources, involvement in other sporting activities, physical growth spurts, or injuries and health problems. The goal is to account for the different paths and rates of improvement of performance of different athletes by examining the effects of individual differences in quality and quantity of practice and training once the effects of other factors, such as injuries and training resources, are controlled.

Our analysis of the retest reliability and validity of these types of reported information has general implications. Any type of interview with athletes concerning their experience needs to consider not only what the researchers want to know, but also what the athletes are able to report accurately. We attempted to design our interview to help the athletes recall actual events and memories from their developmental history to elicit actual memory reports rather than inferences and reconstructions. Our efforts to elicit reliable and verifiable information about perceived characteristics of past engagement in training activities, such as enjoyment and the demand for concentration and physical effort, were relatively unsuccessful. On the other hand, athletes were able to accurately recall many aspects of their development even after decades had elapsed. More generally, we recommend that researchers limit their research to those aspects of the developmental history of athletes that have been shown to be valid indicators of concurrent behavior and can be assessed by verifiable reliable retrospective reports. More specifically, our interview procedure and its valid reported information should provide theoretical and practical insights into the role of various activities in athletes' development in sport. The interview procedure presented in this article allows researchers to highlight the changing environment of athletes throughout their development in sport and has many practical implications for the design of sport programs. The application of this interview procedure with athletes of various levels of performance will provide useful insights to sport consultants regarding important developmental issues such as, the choice of learning objectives, curriculum sequence, and teaching methods at various ages of an athlete's involvement in sport.

REFERENCES

Abernethy, B., Côté, J., & Baker, J. (2002). *Expert decision making in team sport.* Report to the Australian Sports Commission. Brisbane, Australia: University of Queensland.

Bahrick, H. P. (1984). Semantic memory content in permastore: 50 years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113,* 1–29.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104,* 54–75.

Bahrick, H. P., Hall, L. K., & Berger, S. A. (1996). Accuracy and distortion in memory for high school grades. *Psychological Science, 7,* 265–271.

Bahrick, H. P., Hall, L. K., Goggin, J. P., Bahrick, L. E., & Berger, S. A. (1994). Fifty years of maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General, 113,* 264–283.

Baker, J., Côté, J., & Abernethy, B. (2003a). Sport-specific practice and the development of expert decision-making in team ball sports. *Journal of Applied Sport Psychology, 15,* 12–25.

Baker, J., Côté, J. & Abernethy, B. (2003b). Learning from the experts: Practice activities of expert decision-makers in sport. *Research Quarterly for Exercise and Sport*.

Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology, 36,* 253–263.

Beamer, M., Côté, J., & Ericsson, K. A. (1999). A comparison between international and provincial level gymnasts in their pursuit of sport expertise. *Proceedings of the 10th European Congress of Sport Psychology*, Prague, Czech Republic.

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130,* 701–725.

Bloom, B. S. (Ed.). (1985). *Developing talent in young people.* New York: Ballantine.

Borg, G., & Ottoson, D. (Eds.). (1986). *The perception of exertion in physical work*. London: MacMillan.

Bradburn, N. M. (2000). Temporal representation and event dating. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 49–62). Mahwah, NJ: Erlbaum.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5,* 73–99.

Conway, M. (1995). *Flashbulb memories.* Hillsdale, NJ: Erlbaum.

Côté, J. (1999). The influence of the family in the development of talent in sport. *The Sport Psychologist, 13,* 395 – 417.

Côté, J., Baker, J., & Abernethy, B. (2003). From play to practice: A developmental framework for the acquisition of expertise in team sport. In J. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 89–114). Champaign, IL: Human Kinetics.

Côté, J., & Hay, J. (2002). Children's involvement in sport: A developmental perspective. In J. M. Silva & D. Stevens (Eds.), *Psychological foundations of sport* (pp. 484–502). Boston: Allyn and Bacon.

Deakin, J., & Cobley, S. (2003). A search for deliberate practice: An examination of the practice environments in figure skating and volleyball. In J. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 115–136). Champaign, IL: Human Kinetics.

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99,* 550–553.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts, sciences, sports and games* (pp. 1–50). Mahwah, NJ: Erlbaum.

Ericsson, K. A. (2001). The path to expert golf performance: Insights from the masters on how to improve performance by deliberate practice. In P. R. Thomas (Ed.), *Optimising performance in golf* (pp. 1–57). Brisbane, Australia: Australian Academic Press.

Ericsson, K. A. (2003). The development of elite performance and deliberate practice: An update from the perspective of the expert-performance approach. In J. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 49–81). Champaign, IL: Human Kinetics.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence on maximal adaptations on task constraints. *Annual Review of Psychology, 47,* 273–305.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–251.

Ericsson, K. A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: The MIT Press.

Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Towards a general theory of expertise: Prospects and limits* (pp. 1–29). Cambridge, UK: Cambridge University Press.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327–358.

Friedenreich, C. M., Courneya, K. S., & Bryant, H. E. (1998). The lifetime total physical activity questionnaire: Development and reliability. *Medicine & Science in Sports & Exercise, 30,* 266–274.

Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport and Exercise Psychology, 20,*12–34.

Hodge, T., & Deakin, J. (1998). Deliberate practice and expertise in the martial arts: The role of context in motor recall. *Journal of Sport and Exercise Psychology, 20,* 260–279.

Hodges, N. J., & Starkes, J. L. (1996). Wrestling with the nature of expertise: A sport specific test of Ericsson, Krampe, and Tesch-Römer's (1993) theory of "deliberate practice". *International Journal of Sport Psychology, 27,* 400–424.

Hue, O., Le Gallais, D., Chollet, D., & Prefaut, C. (2000). Ventilatory threshold and maximal oxygen uptake in present triathletes. *Canadian Journal of Applied Physiology, 25,* 102–113.

Hume, P. A., Hopkins, W. G., Robinson, D. M., Robinson, S. M., & Hollings, S. C. (1994). Predictors of attainment in rhythmic sportive gymnastics. *The Journal of Sports Medicine and Physical Fitness, 33*(4), 367–377.

Kihlstrom, J. F., Eich, E., Sandbrand, D., & Tobias, B. A. (2000). Emotion and memory: Implications for self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 81–100). Mahwah, NJ: Erlbaum.

Krampe, R. Th., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General, 125,* 331–359.

Markus, G. B. (1986). Stability and change in political attitudes: Observed, recalled and "explained." *Political Behavior, 8,* 21–44.

Massey, D. S. (2000). When Surveys fail: An alternative for data collection. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 145–160). Mahwah, NJ: Erlbaum.

Menon, G., & Yorkston, E. A. (2000). The Use of memory and contextual cues in the formation of behavioral frequency judgments. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 63–80). Mahwah, NJ: Erlbaum.

Noble, B. J., & Robertson, R. J. (1996). *Perceived exertion*. Champaign, IL: Human Kinetics.

Ropponen, A., Levalahti, E., Simonen, R., Videman, T., & Battié, M. C. (2001). Repeatability of lifetime exercise reporting. *Scandinavian Journal of Medicine & Science in Sports, 11,* 185–192.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96,* 341–357.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103,* 736–760.

Salmela, J. H. (1995). Learning from the development of expert coaches. *Coaching and Sport Science Journal, 2,* 3–13.

Schmidt, H. G., Peeck, V. H., Paas, F., & van Breukelen, G. J. P. (2000). Remembering the street names of one's childhood neighborhood: A study of very long-term retention. *Memory, 8,* 37–49.

Sedgwick, W. A., Ericsson, K. A., Beamer, M., & Côté, J. (1999, September). *An assessement of the training variables of international and national level rowers.* Paper presented at the meeting of the Association for the Advancement of Applied Sport Psychology, Banff, Canada.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Singer, R. N., & Janelle, C. M. (1999). Determining sport expertise: From genes to supremes. *International Journal of Sport Psychology, 30,* 117–150.

Smith, J. C., & Crabbe, J. B. (2000). Emotion and exercise. *International Journal of Sport Psychology, 31,* 156–174.

Soberlak, P. & Côté, J. (2003). The developmental activities of professional ice hockey players. *Journal of Applied Sport Psychology, 15,* 41–49.

Starkes, J. L. (2000). The road to expertise: Is practice the only determinant? *International Journal of Sport Psychology, 31,* 431–451.

Starkes, J. L., Deakin, J. M., Allard, F., Hodges, N. J., & Hayes, A. (1996). Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 81–106). Hillsdale, NJ: Erlbaum.

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59,* 291–297.

Weston, A. T., Petosa, R., & Pate, R. R. (1997). Validation of an instrument for measurement of physical activity in youth. *Medicine & Science in Sports & Exercise, 29,* 138–143.