Developing Rigor in Qualitative Research

Smith, Brett; McGannon, Kerry R

DOI:
10.1080/1750984X.2017.1317357

Developing Rigor in Qualitative Research: Problems and Opportunities within Sport and Exercise Psychology

Abstract

Qualitative research has grown within sport and exercise psychology and is now widely conducted. The purpose of this review is to discuss three commonly used ways to demonstrate rigor when conducting or judging qualitative research in sport and exercise psychology. These are the method of member checking, the method of inter-rater reliability, and the notion of universal criteria. Problems with each method are first highlighted. Member checking and inter-rater reliability are shown to be ineffective for verification, trustworthiness, or reliability purposes. Next, universal criteria within the context of Tracy's (2010) heavily drawn on paper within sport and exercise psychology is problematized. Throughout the discussion of each method and universal criteria more suitable possibilities for conducting rigorous qualitative research are offered. The paper concludes that to support high quality qualitative research, scholars - including journal editors and reviewers - need to change how rigor is developed and judged, rather than perpetuate the problems with how it has been commonly evaluated in the past. Recommendations for developing rigor when conducting and/or judging qualitative research within sport and exercise psychology are also offered.

Key Words: Member Checking; Inter-rater Reliability; Universal Criteria; Research Quality

Qualitative research has been utilized as a form of inquiry within sport and exercise psychology for over three decades. During this time there has also been a rapid growth of qualitative research in the field. For example, in their up-dated review of qualitative research in three North American journals (i.e., *Journal of Applied Sport Psychology*, *Journal of Sport and Exercise Psychology*, and *The Sport Psychologist*) Culver, Gilbert and Sparkes (2012) highlighted that between 2000–2009 there was a 68% increase in the percentage of qualitative studies published since the 1990's (from 17.3% to 29%). A significant increase in the number of different authors publishing qualitative research in these journals was also noted.

In 2009 when Culver et al's. (2012) review period ended the international journal *Qualitative Research in Sport, Exercise and Health* began. Attracting hundreds of submissions yearly, and with 5 issues per year, the journal has published empirical papers within sport and exercise psychology, supported different paradigms and theories, and encouraged innovative methods and methodologies. In recent years other international journals (e.g., *Psychology of Sport and Exercise* and *International Journal of Sport and Exercise Psychology*) have similarly published different kinds of qualitative research, creating space for work grounded in different and/or multiple methods, methodologies and ways of knowing. The growth of qualitative research within the broad field of sport and exercise is further evident in the increasing number of heavily cited books devoted solely to qualitative research (e.g., Jones, Brown, & Holloway, 2012; Pitney & Parker, 2009; Smith & Sparkes, 2016a; Sparkes & Smith, 2013; Young & Atkinson, 2012). Moreover, a growing number of conferences and workshops are attempting to address the demand for qualitative research from students, researchers, practitioners, and policy makers. For instance, the bi-annual *International Conference for Qualitative Sport and Exercise* (see twitter @QRSE2018) and quarterly congress by *The European Federation of Sport Psychology* (see

http://www.fepsac.com) have hosted workshops dedicated to qualitative research and showcased hundreds of qualitative research papers from established scholars and newcomers from around the world.

In light of the foregoing, it would appear that within sport and exercise psychology qualitative research has flourished considerably in recent years. Yet despite flourishing, we cannot be complacent. Like any vibrant field, important intellectual developments within qualitative research have taken place. These developments include work on rigor. Keeping abreast of intellectual developments is of course crucial. Developments in our thinking can mean that certain historically popular qualitative methods and methodologies might now need rejecting, corrective action, or exigent deliberation. Thus, in order for high-quality research to be conducted researchers need to stay engaged with contemporary methodological thinking by, for example, connecting with recently published work (e.g., Birt, Scott, Cavers, Campbell, & Walter, 2016; Burke, 2016; Levitt, Motulsky, Wertz, Morrow, & Ponterrotto, 2016; Morse, 2016) on matters like rigor. When we ignore such thinking there is a risk of producing outdated, flawed, stagnant and/or limited research.

As the qualitative research landscape continues to rapidly expand and flourish, engaging with contemporary literature concerning the latest thoughts and developments can be challenging. For instance, given our investments in certain methods or methodologies it can be difficult to read work that substantially questions a method or methodology used often in the past. With the large amount of research being regularly published, it may also be difficult to keep abreast of developments in the qualitative methodological literature, even when one is interested in embracing them. In that regard, reviews can be useful resources to take stock of developments, offering ways forward in light of said developments. The purpose of this review paper is to discuss one contemporary development around the theme of ‘rigor in qualitative research’. Organized under that theme, three widely used ways of

demonstrating rigor in qualitative research within sport and exercise psychology are critically attended to. These are the method of member checking, the method of inter-rater reliability, and the notion of universal criteria within the context of Tracy's (2010) heavily drawn upon and/or cited paper in the field. Numerous problems with each way of approaching rigor are highlighted. Throughout discussing each approach to rigor, constructive possibilities for conducting quality qualitative research are offered. We conclude with a set of recommendations for considering rigor when conducting or judging qualitative research within sport and exercise psychology.

In accomplishing the above purpose, we do not single out specific papers within sport and exercise psychology that have utilized the methods of member checking, inter-rater reliability and/or universal criteria outlined and problematized. Singling out work might be counterproductive to our hope that the points brought forward may be read with intellectual curiosity and an openness to consider each point made. We therefore invite those conducting qualitative research in sport and exercise psychology, or are considering doing it, to engage with each point in order to develop informed and reflexive decisions about rigor. Editors and reviewers of journals are also encouraged to consider the points raised, including the recommendations for developing rigor in qualitative research that close the paper.

Rigor in Qualitative Research: Problems and Possibilities

The notion of rigor is often viewed as a necessary marker of quality by researchers, reviewers, journal editors, and research panel members. As Tracy (2010) argued, for "qualitative research to be of high quality, it *must* be rigorous" (p. 841). However, what is meant by rigor can vary amongst scholars; it can mean different things to different people. Such meanings may include, but are not limited to, the intellectual precision, robustness, appropriateness, sufficiency, and cohesiveness of concepts, methodologies, epistemology, ontology, and methods deployed in the research process and output (Burke, 2016; Saldaña,

2013; Tracy, 2010). With the various meanings and possibilities of rigor recognized, when it comes to qualitative research within sport and exercise psychology, rigor has largely been described as a marker of excellence sought through method. Methods are techniques which, when properly applied, are said to provide rigor.

Member Checking

One extensively used method of rigor within qualitative research is member checking. Member checking was popularized within the qualitative research literature by Lincoln and Guba (1985). Member checks, or what is sometimes also termed ‘respondent or participant validation’, involve the participants of a project assessing the trustworthiness of research in terms of validating the credibility of qualitative data and results. In member checks validation is often done by first returning the data (e.g. interview transcripts) and/or results (e.g. themes and interpretations) to the research participant and then asking them to provide input on whether the data is accurate and/or results accurately reflect their experiences. If the participant confirms the accuracy of the data and/or results, the findings can be deemed credible and the research is valid. Thus it is often suggested, either implicitly or explicitly, that member checks are a means of controlling or correcting the subjective bias from the researcher and/or a useful means of checking the truth of any knowledge (Birt et al., 2016; Lincoln & Guba, 1995). Framed in this manner, member checking is denoted as providing a check that a researcher has made contact, however subtle or approximate, with the social reality independent of their interest in, or knowledge of reality, thereby enabling the adjudicating between trustworthy and untrustworthy interpretations (Birt et al., 2016; Lincoln & Guba, 1985).

Within sport and exercise psychology member checking is frequently used when conducting qualitative research. As Culver et al. (2012) found, member checking rose from being used in 21% in qualitative studies in the 1990’s to 50.3% in 2000-2009. Such a

statistic, and the frequency in which one member checking is still being used in sport and exercise psychology research, is however troubling. That is because there are major problems with member checking as a form of quality control and a benchmark of rigor when it comes to verifying qualitative research. Before highlighting evidence and practical problems in this regard, the epistemological and then ontological problems of member checking need emphasizing. An emphasis on epistemology and ontology is necessary because all methods are informed, either knowingly or unknowingly, by an epistemology (i.e. what is the relationship between the inquirer and the known? What types of knowledge might be legitimately known?) and ontology (i.e. what is the nature of reality?) (Eakin, 2016; Giardina, 2017; McGannon & Schweinbenz, 2011; Sparkes & Smith, 2014). Put differently, research methods cannot be divorced from their philosophical undercarriage. Thus, echoing Braun and Clarke (2013) on member checks, Birt et al. (2016) confirm that "member checking should not be considered merely as a simple technical step in any study" (p. 1810). Member checking is an intellectual process which, they note, presents distinct epistemological and ontological challenges. Such challenges are especially pertinent to producing rigorous qualitative research when rigor is considered not only in terms of method but also as the intellectual precision, robustness, appropriateness, and cohesiveness of the epistemology and ontology utilized (Morse, 2016).

Within sport and exercise psychology, Lincoln and Guba (1985) are often cited to support the use of member checking as a useful, if not necessary, method of quality control in qualitative research. Indeed, scholars employing member checking have proposed that this method was the most crucial technique for achieving credibility (Culver et al., 2012). Although Lincoln and Guba accepted a world of multiple and mind-dependent realities (i.e. ontological relativism), they also discussed member checking as informed by epistemological foundationalism. Epistemological foundationalism refers to the assumption that method, like

member checking, itself is neutral and can thereby control for bias and be the repository of procedural objectivity to sort out the more or most trustworthy from the less trustworthy. However, a major problem with epistemological foundationalism is that methods are not neutral, objective, or unbiased but rather are dependent on people (Culver et al., 2012; Smith & Deemer, 2000). In the case of member checking, it *is* the researcher and participant who are the member checkers. Moreover, like all people, researchers and participants alike are unable to step outside of their own experiences and history or rise above and separate themselves from the study of the social world (Denzin, 2017).

Thus, as various scholars (e.g., Braun & Clarke, 2013; Denzin, 2017; Smith & Deemer, 2000) remind us, there is no possibility of producing theory-free knowledge when using methods like member checking. The researcher and participant, no matter how hard they try, will always influence the method and any knowledge claim that follows from using it will be infused with their subjectivities (McGannon & Smith, 2015; Smith & Deemer, 2000). As Braun and Clarke (2013) noted when talking about member checks, researchers cannot simply represent experience. Understanding and representing peoples' experiences requires "interpretive activity; this is always informed by our own assumptions, values and commitments" (p. 285). Given that point, member checking cannot deliver objective knowledge. Nor can it provide an independent foundation to adjudicate valid research from less valid research. As such, any claim that the rigor of research has been enhanced through member checking cannot be made good and demonstrated. It is simply a claim – and a problematic claim - because it does not have a *necessary* epistemic foundation to support it.

In light of such problems, and demonstrating a scholarly openness to re-consider presuppositions and change, Guba and Lincoln (1989) and Lincoln and Guba (2000) later abandoned a foundational epistemology (see also Guba & Lincoln, 2005; Lincoln, 1995; 2010). Of course, such reworking of these ideas raises the question as to why sport and

exercise psychology researchers still use and cite the 1985 work by Lincoln and Guba, which these authors themselves acknowledged was deeply problematic. But a more productive and useful question to ask is how might researchers now proceed epistemologically and ontologically with member checking when it has been problematized? One common response within sport and exercise psychology, either implicitly or explicitly, is to connect with both epistemological constructionism and ontological realism. The former accepts that there cannot be theory-free knowledge because a person's understanding of reality is only known through their experiences (i.e., knowledge is socially constructed and thus fallible). The latter claims that there is a reality independent of us that, however subtle or approximate, can be known. The combination of epistemological constructionism and ontological realism informs what has variously been referred to within sport and exercise psychology as neo-realism, subtle realism, post-positivism, or quasi-foundationalism (Smith & Sparkes, 2016a).

Despite the appeal of both epistemological constructionism and ontological realism, there is a major problem with a dual commitment to that epistemology and ontology which has important implications for member checking. The problem starts with a commitment to an epistemology and ontology that is incompatible and, in turn, untenable in terms of holding both together simultaneously. In other words, combining epistemological constructionism and ontological realism is neither possible nor sustainable - the two don't fit together. That is because, on the one hand, committing to the belief that knowledge is socially constructed means that theory-free knowledge is *unachievable*. On the other hand, believing that there is a social reality independent of us that can be discovered - however ideal, approximate or subtle - means that theory-free knowledge *can be achieved*. Thus, the realist ontology held by the researcher contradicts the constructionist epistemological they hold - they cannot have it both ways.

Given the incompatibility of epistemological constructionism and ontological realism,

researchers must move in one of two directions. The first direction is to drop ontological realism and along with that, what must now be accepted is epistemological constructionism and also ontological relativism (i.e. multiple and mind-dependent realities)[1]. If researchers wish to hold onto ontological realism then the second direction is to drop epistemological constructionism and confirm the existence of foundations and of a reality outside of ourselves that can be known independently/objectively through the appropriate use of techniques. And, as Smith and Deemer (2000), and Smith and Hodkinson (2009) emphasize, this assumption (the reality) has to be cashed in - demonstrated - to do its work of adjudication by sorting out the trustworthy interpretations from untrustworthy ones. The move to retain ontological realism has however a major problem. As Smith and Hodkinson (2009) argued, one can assume all one wants about an independent social reality, but the

> problem is that there is no way to "get at" that reality as it really is. And, if one cannot capture that social reality as it really is, then that reality cannot be called on to do the adjudicate-the-different-claims-to-knowledge work asked of it. This is the whole problematic posed by the idea that no matter how hard we try, we cannot achieve theory-free observation or knowledge. (p. 34)

Accordingly, researchers can talk all they want about an independent social reality, accurately depicting people's reality, or producing credible or trustworthy results that correspond to the reality. However, unless they can somehow demonstrate through a method like member checking how they can overturn the long standing and widely argued point (see Smith & Deemer, 2000) that theory-free knowledge cannot be achieved, then accessing an independent social reality, depicting people's reality accurately with certitude, or producing credible or trustworthy results that correspond to the reality is untenable and a chimera. Ultimately, as Smith and Hodkinson (2009) argued, researchers are unable to establish how through method an external social reality can be objectively accessed and known as it really

is such that it can do the adjudication work required of it.

Of course, a researcher might still dispute the arguments above and persist in claiming that member checking is useful to ensure rigor in qualitative research. However, there are at least three reasons why that method is problematic for ensuring rigorous research. As noted already, the first of these reasons is that researchers have been unable to show how to make contact with the external referent point – the reality – to which they *must* appeal in order to give member checking a standing beyond the socially constructed judgments researchers and participants make (Smith & Hodkinson, 2009). That is, if a researcher claims that methods like member checking can help sort out the trustworthy from the untrustworthy interpretations, it is a necessity for them to demonstrate how that method can make, and has made, contact with the reality independent of them. Yet, contacting that reality cannot be done because a researcher is unable to override their human finitude to achieve theory-free knowledge. Thus, without being able to do the necessary work of making contact, however subtle or approximate, with the independent social reality a researcher cannot objectively sort out the trustworthy from the untrustworthy interpretations through member checking and, subsequently, that method remains ineffective for enhancing rigor.

Secondly, despite *claims* about the value of member checking, in a recent review of published literature by Thomas (2017) it was concluded that there is “no evidence that routine member checks enhance the credibility or trustworthiness of qualitative research” (p. 37). Thirdly, any claim that member checking can verify or add rigor falls further apart because of the numerous insurmountable *practical problems* researchers face when using this method as a form of verification and quality control.

One practical problem researchers may encounter is the possibility that the participant and researcher might provide interpretations of the findings that contradict each other. However, member checking provides no means to decide between contradictory claims to

knowledge based on experience (Sparkes & Smith, 2014). Member checking is also problematic as a verification method because a researcher is unable to know with certainty that each participant has faithfully engaged with member checking. For instance, when faced with a large data set and findings a participant might briefly skim over what was offered. To save time, and perhaps to even appease the researcher without him or her knowing, the feedback from the participant may end up like this: 'Yes, I agree. Everything is 100% correct'. At this point, it should equally be noted that in sport and exercise psychology it is nigh on impossible to find a paper that discloses any disagreements made by a participant during the process of member checking. It is as if all our many studies using that method, totaling hundreds if not thousands of participants, produce 100% member agreement. Some might argue this agreement supports the success of member checking as a method for sorting out trustworthy from untrustworthy interpretations. However, given member checking cannot ensure rigor for reasons highlighted above this then begs the question as to why we never hear about disagreements.

Another practical problem with member checking relates to power relations within research. For example, when asked to member check findings a participant might not comprehend the results. Yet, in order to present themselves as an 'equal' to the researcher they might choose to agree with the findings. Alternatively, as Estroff (1995) cautioned, a participant may defer to the researcher as the 'expert', accepting all they say even if they have different viewpoints about the data or findings. Given such potential problems a researcher might put in place strategies to minimize power differentials during the research. Conversely, later they might delicately ask the participant if they are simply agreeing or deferring to them when conducting member checks. However, ultimately power relations never can be truly eradicated and thus the researcher can never know with certainty if the participant has simply agreed with them or deferred to their 'expertise'.

Moreover, there is the practical problem related to time. Although researchers unfortunately rarely report on how long after data collection member checking occurred, given the length of time to transcribe/record and analyze it is likely that member checks occur some time after data has been collected. This length of time might be much longer when a researcher collects longitudinal data using different methods (e.g., multiple mobile interviews, visual methods, observations, diaries, and digital methods). Data and interpretations might also be more complex and layered than when gathered from interviewing people just once for a short amount of time (Smith & Sparkes, 2016c). In such cases asking the participant to consider the accuracy of statements or the interpretations offered may pose a problem for member checks. For instance, the person may no longer find that their experiences 'ring true' or align with previous experiences. Recall bias or memory distortion is not however the issue here. That is because memories are partly socially constructed, people's perceptions change and contradict, and our discourses constitute and become experience rather than transparently reflecting experiences[2]. The issue is that member checking is influenced by time which, in turn, can limit the effectiveness of that method when one seeks to use it to demonstrate accuracy or get at the truth and accuracy of experience.

A further reason why member checking is ineffective as a verification method is that people's political leanings and personal interests may, consciously or unconsciously, influence member checking. For instance, a participant may consent to member checks because they have a political and personal investment in the outcomes of research. When member checks are conducted the participant may, unbeknownst to the researcher, reject or censor the findings if these cast them in a poor light. Equally, without the researcher aware, participants may reject a finding if it questions or conflicts with personal interests. For instance, a person who has helped create a physical activity policy on tackling obesity might

disagree with a researcher's critique of it because it conflicts with their own personal interests in ensuring, as part of their job and future employment options, that the policy is taken-up by government. When one is aware of such possibilities a researcher might again put strategies in place to try to help constrain the political and personal investments that impact on the accuracy of member checking. However, because it is impossible to achieve theory-free knowledge and know for certain that an external reality has been found to do the necessary work of adjudicating if political or personal interests have been truly constrained, a researcher is unable to know with absolute certainty if interests have influenced member checking. Once again, member checks cannot then be cashed in on, or relied upon, to support any claim that the method has enhanced research quality and rigor.

In light of all the various problems highlighted, and with no way to overcome these nor any evidence to support that member checks work as a form of validating truth claims, it would appear that member checking is an ineffective method for the purposes of verification. Thus, as Morse (2016) recently noted, as a technique for achieving validity or reliability "member checking is not recommended" (p. 1216). Unless researchers can demonstrate otherwise, it is perhaps time that sport and exercise psychology researchers give up on the illusion that member checking can help verify research and act as a benchmark of quality control and rigor. Where might then this leave a researcher in terms of enhancing rigor? One possibility is to reframe member checking as member reflections (Braun & Clarke, 2013; Tracy, 2010).

Member reflections is not about verifying results, finding correspondence with the truth, or getting at the independent reality. Rather, one aim of member reflections is to generate additional data and insight. For instance, together a researcher and participant might engage in member reflections to explore any gaps in the results or similarities they share concerning interpretations of the findings (Schinke, Smith & McGannon, 2013). Moreover,

instead of framing member checks as a method to sort out between contradictory or different claims to knowledge, researchers might more productively reframe it as a *practical opportunity* to acknowledge and/or explore *with* participants the existence of contradictions and differences in knowing. As part of this co-participatory process and dialogue, participants and researchers might both consider such questions as: What might be done when we - the researcher and participant - disagree over the findings? Should the interpretations of the participant be privileged over the researchers' (or visa versa)? Or can the differences be somehow incorporated in the final report? There is no universally agreed, final, or singularly right answer to such questions. Instead, different responses may be given as each is applied contextually, to actual practice and different projects as situations demands (Schinke et al., 2013).

In addition to reframing the aim of member checking as an opportunity to help work with participants and facilitate the inclusion of complementary or contradictory results so that a meticulous, robust, and intellectually enriched understanding of the research might be further developed, member checks may be useful for ethical reasons within the research process. We stress 'may' because this method can engender ethical problems. For instance, when researchers decide to share findings with participants, disappointment, hurt feelings, and embarrassment may follow for them both (Sparkes & Smith, 2014). The participants could be unhappy with the analysis, feel that they are being depicted insensitively, or perceive that the researcher has unfairly used their power to expose vulnerabilities. In these circumstances, researchers can believe and experience that their ethical commitment to do no harm has been violated. With such caveats in mind, member checking might be of use as part of adopting a culturally responsive relational reflexive ethics position (McGannon & Smith, 2015; Palmer, 2016). For example, as part of adopting that ethical position, promoting dignity, mutual respect and connectedness between the researcher and participant may be

enhanced through the sharing of findings and experiences. When couched within the notion and practice of a culturally responsive relational reflexive ethics position, member checking may moreover be a valuable and useful tool when dealing with the possibility of breaching confidentiality via deductive disclosure (i.e. when traits and rich descriptions of people in research reports unintentionally reveal who they are) (see Sparkes & Smith, 2014).

Inter-rater Reliability

According to Culver et al. (2012), in sport and exercise psychology 82.2% of qualitative studies clearly reported reliability testing. The most extensively used reliability technique within the field is *inter-rater reliability*. Sometimes termed 'investigator triangulation', inter-rater reliability is a method that aims to ensure results are reliable in the sense of being reproducible and consistent by employing intercoder reliability and intercoder agreement (Campbell, Quincy, Osserman, & Pedersen, 2013). The former requires two or more capable researchers operating in isolation from each to independently code data without negotiation. The latter requires that two or more researchers come together to compare codes and then reconcile through discussion whatever coding discrepancies they may have for the same unit of text. When a high level of agreement/consensus is demonstrated at the end of this process, the coding is deemed reliable. Hence, inter-rater reliability is about two or more researchers independently coding data and coming to an agreement over the codes to check that coding is replicable (Lincoln & Guba, 1985). The use of inter-rater reliability is typically written up in the methods section as follows: "To ensure reliability and avoid bias, three trained researchers analyzed the data independently. Following a discussion over disagreements, there was 87% consensus on the codes (or themes)".

Despite the appeal and wide use of inter-rater reliability as described above, this method is *ineffective* for helping to ensure reliable qualitative research. There are various reasons as to why inter-rater reliability as traditionally used does not work as a form of rigor

in qualitative research. One reason concerns the problem involving unitization (i.e., being able to identify appropriate blocks of text for a particular code or codes). Unitizing the text, according to Campbell et al. (2013), is “a problem insofar as different coders may unitize the same text differently. Why? Because they may disagree on which segments of text contain a particular meaning” (p. 302). For instance, as Campbell et al. (2013) note, although two coders might each identify a string of text for the same code with each of the strings overlapping, they might “still vary in length because one coder includes text providing background information that helps establish the context for the code in question but the other coder does not” (p. 302). As such, identifying the appropriate unit of analysis can be difficult which “makes it hard to assess intercoder reliability and agreement” (p. 302).

A further problem can be traced back to the issue that was raised earlier concerning member checking, which is that human beings - who are the coders - cannot, no matter how hard we try, produce theory-free knowledge (Guba & Lincoln, 2005; Smith & Deemer, 2000). Thus, the theory, hypothesis, framework, or background knowledge held by the researcher inescapably influences coding and is inextricably linked with the research process. Consider the following example (see also Communication Studies 298, 2000; Madill, Jordan & Shirley, 2000). Two researchers are chosen to code data from two different universities. One is a cognitive exercise psychologist who, informed by self-determination theory, thinks that people have an innate self with certain natural needs. In contrast, the other researcher is a discursive exercise psychologist who, informed by social constructionism, believes that selves are socially constructed through discourses and particular ways in which language is used during interaction with one’s self and others. Now imagine the two researchers independently code the same data in relation to ‘the self’ (as conceptualized by the underlying theoretical perspective to which each researcher subscribes). Partly because they are influenced by very different and incompatible theoretical ideas, it is very unlikely that the

researchers would code similarly or agree after discussions over the meaning of the codes. Add a third person to the mix, who adheres to social identity theory or feminist post-structuralism, and more disagreements and different ‘results’ in terms of coding would probably arise. Conversely, when researchers who hold similar theoretical interests come together their shared background knowledge and theoretical lens influences the coding, leading perhaps to many agreements and fewer disagreements. Accordingly, because we can never achieve theory-free knowledge, inter-rater reliability will always be influenced by people’s theoretical proclivities, thereby making the method (including results) related and tied to people’s subjectivities and histories within academia.

Likewise in terms of the relationality of researchers to knowledge and the research process, it is important to raise the question ‘who did the coding?’ and ‘who discussed/compared codes’ to establish inter-rater reliability. A close reading of published sport and exercise psychology research will often reveal that the coders are PhD student(s)/research assistant(s) and a supervisor/grant lead. This order of authorship is often understandable given the investments each has in the research and certain power hierarchies in research mentoring and institutions. However, there is nothing ‘independent’ or ‘objective’ about coming together to discuss codes and calculating a certain percentage of agreement to indicate rigor. Power differentials and gender dynamics between coders, plus the age, nationality, past training, and so on of each, can strongly influence inter-coder agreement. For example, disagreements between coders might occur during initial discussions over the codes. But, these disagreements might be quickly ‘resolved’ due to the power differences that operate, implicitly or explicitly, between the coders. As Campbell et al. (2013) recognized, when discussing codes a student/assistant might defer to their supervisor/grant lead due to intimidation or a history of them removing ‘dissenters’ from the group. There equally could be an implicit threshold of disagreement where the student/assistant feels that she or he

should only dissent or disagree a few times and/or not in particularly assertive or forceful ways. Or, following discussing coding disagreements, a student/assistant might defer to the supervisor/grant leader because of a deep respect they hold for them and their knowledge of the subject matter.

Whatever the case might be for these power differentials and associated practices/behaviours, the point is again that coding is done by humans who are not only intimately a part of any understanding we have of what counts as knowledge, but also relational beings (Gergen, 1999). The inevitable result of this relationality is that inter-rater reliability will always be influenced, no matter how hard we try to erase of control from them, by the background of the researchers plus the power, age, and gendered relations that operate between them. Thus, inter-rater reliability cannot be viewed, nor used, as an unbiased method or produce independent final results, leaving anyone that seeks to use it to ensure their qualitative research is reliable with major problems that cannot entirely be overcome because theory-free knowledge is impossible.

Another problem with inter-rater reliability is that it always possible that coders might agree occasionally by chance. Also problematic is that there is no agreed upon threshold in the literature for what constitutes a numerically satisfactory level of agreement among coders to achieve reliability and more or less rigor. Is it 78%, 87%, or 90% agreement? An examination of papers in sport and exercise psychology will reveal that what passes for an acceptable level of intercoder reliability varies considerably according to the standards of different researchers as well as the method of calculation (Culver et al., 2012). How then does one truly know what is acceptable for qualitative research to be deemed reliable? Is a study that has 90% intercoder agreement ‘better’ than one that has an 87% or 78% agreement? Or is it that simply reporting a ‘high’ figure is a marker of acceptability? Will that simply do? Our point is that without an objectively arrived at threshold for what

constitutes a numerically satisfactory level of agreement such questions cannot be answered with any certainty. Thus, researchers, reviewers and journals editors have no foundation to independently assess reliability claims. With no universally agreed threshold severe doubts are again cast on inter-rater reliability.

As noted, between 2000 and 2009 Culver et al. (2012) found that in three sport and exercise psychology journals 82.2% of qualitative studies used a reliability test. A close look at the qualitative literature in the field over recent years will reveal that inter-rater reliability is still often used. That 82.2% statistic and continued use of inter-rater reliability is troubling given the many problems of the method. Moreover, the numerous problems associated with inter-rater reliability have led researchers to conclude that it is a myth (Morse, 1997), a misconception (Cook, 2011), inappropriate for interpretive qualitative research (Levitt et al., 2016; Madill et al. 2000), not worth pursing (Braun & Clarke, 2013), and a flimsily retrofitted procedure in qualitative clothes to be avoided (Eakin, 2016). Accordingly, because this technique as traditionally conceived and widely used in sport and exercise psychology cannot effectively test if qualitative research is reliable, it should be dropped for reliability and rigor purposes.

If inter-rater reliability as traditionally used is discarded as an indicator of reliability and rigor, where then does this leave sport and exercise psychology researchers? Let us offer two possibilities. First, for researchers who still view reliability a concern when doing qualitative research, additional and different ways of doing inter-rater reliability may be adopted. One option might include adopting the guidelines described by MacPhail, Khoza, Abler and Ranganathan (2016), which seek to establish intercoder reliability in qualitative research. These guidelines involve deductively developing a coding frame and set of firm coding rules (e.g. about when a chunk of text starts and ends and the degree of specificity to be adopted when coding) prior to a team independently coding of data. The finalized code

frame and rules are then used to guide the subsequent coding process. After the team come together to compare codes a statistical calculation (i.e. Cohen's kappa) is utilized to measure reliability among the team of coders. Such guidelines align with the assumptions of post-positivism, quasi-foundationalism, and neo or critical–realism. Although using such guidelines are not without challenges, have only been trialed with semi-structured interview data, may constrain creativity, can limit the identification of unanticipated yet insightful knowledge, and risks producing superficial qualitative research (Morse, 1997, 2016), for those who view reliability as a purview of qualitative research then new developments like these must be considered. The consideration of such developments is especially required in light of the numerous inescapable problems outlined that go with how inter-rater reliability has traditionally been done within sport and exercise psychology.

Another possibility, and one that aligns with interpretive forms of qualitative research that commit to epistemological constructionism and ontological relativism, is to reject reliability as an appropriate criterion for judging the rigor of qualitative research. Numerous researchers (e.g., Braun & Clarke, 2013; Cook, 2013; Cosh, LeCouteur, Crabb, & Kettler, 2013; McGannon & Spence, 2010; Pitney & Parker, 2009; Sparkes & Smith, 2014; Wolcott, 1995) have taken this option for very good reasons. The problems with inter-rater reliability already highlighted are important reasons as to why reliability is rejected as a criteria for evaluating the rigor of qualitative research. Another reason is that applying reliability criteria to qualitative research is incompatible with the belief that theory free knowledge is unachievable and that realities are subjective, multiple, changing, and mind-dependent. It should however be made clear that for those that subscribe to ontological relativism and epistemological constructionism our inability to produce theory free knowledge and get at the social reality independent of us is not the problem; it is simply a recognition that we are finite human beings (Smith & Deemer, 2000).

Furthermore, a lack of concern about reliability, and importantly not apologizing for this lack of concern (Wolcott, 1995), is justified since reliability doesn't make sense when collecting qualitative data. As Pitney and Parker (2009) acknowledge, reliability assessments do not fit with the assumptions of the qualitative research enterprise in terms of collecting data. For them, because qualitative researchers do not conduct the same interview twice, "they do not need to worry about whether data can be reproduced" (p. 62). For instance, seeking reliability becomes nonsensical because a qualitative researcher seeking rich and personally meaningful information from people in interviews does not ask the same questions, in precisely the same order, with the same non-verbal expressions or emotional tone, in repeated social contexts and situations, with no change in their knowledge based on previous interviews, and so on. In short, as Sparkes and Smith (2014) put it, we cannot step into the same stream twice.

Such arguments related to reliability also apply to other forms of qualitative data collection methods. Wolcott (1995) states that reliability remains beyond the pale for research based on observation in the social world.

> In order to achieve reliability in that technical sense, a researcher has to manipulate conditions so that replicability can be assessed. Ordinarily, fieldworkers do not try to make things happen at all, but whatever the circumstances, we cannot *make* them happen twice. When something does happen more than once, we do not for a minute insist that this repetition is exact. (Wolcott, 1995, p. 167)

Another reason why numerous qualitative researchers consider reliability as less than relevant to their concerns, and feel no reason to apologize for it, relates to interpretation. Superficial and thin interpretations raise the chances of agreement among researchers because of the 'level' of interpretation offered. However, good qualitative research seeks to offer complex, layered, and rich interpretive insights of people's lives. Seeking to keep coding

reliable in the conventional sense, or aiming always for agreement over findings, can often come at the expense of that goal (Morse, 2016). As Kvale (1996) put it, seeking reliability can "lead to a tyranny by the lowest common denominator: That an interpretation is only reliable when it can be followed by everyone, a criterion that could lead to a trivialization of the interpretations" (p. 181).

None of the above points are intended to suggest that reliability should be abandoned in *quantitative* work. Reliability makes sense for this work given the assumptions, aims, and, for instance, the importance in determining a psychometrically sound research instrument. However, whilst reliability is needed in quantitative research, it can be considered inappropriate for qualitative inquiry due to the various reasons outlined. It should also be made clear that in opting to view reliability as no longer a concern in an unapologetic manner, rigor is still important for doing high quality qualitative research but with a different meaning as to what constitutes rigor and how it is achieved. Instead of thinking in terms of inter-rater reliability as a means to achieve rigor and quality, rigor can be enhanced through other ways. One way is via 'critical friends', which is a process of critical dialogue between people, with researchers giving voice to their interpretations in relation to other people who listen and offer critical feedback. The role of the critical friends is "not to 'agree' or achieve consensus but rather to encourage reflexivity by challenging each others' construction of knowledge" (Cowan & Taylor, 2016, p. 508). The role is to provide a theoretical sounding board to encourage reflection upon, and exploration of, multiple and alternative explanations and interpretations as these emerged in relation to the data and writing. As Wolcott (1994) put it:

> The crucial element in soliciting feedback seems to be to engage in dialogue about interpretive possibilities. As with writing, engaging in a dialogue requires that you first give voice to your thought processes. In the process of giving voice to your

> thoughts you give access to them as well. There is some subtle reciprocity involved here; you are never totally free to ignore the suggestions of invited critics….Every invited reviewer is…a potential source of insight into the adequacy of your descriptive account, the incisiveness of your analysis, the depth of your interpretation. Every opinion offered is also a reminder that for every additional viewer there is an additional view. (p. 42)

Thus, in contrast to inter-rater reliability conversations, the notion of critical friends acknowledges that while there can be agreement, such agreement does not mean that the truth has been found or that the research is stable or striving toward reproducibility, and achieved rigor as a result. The use of critical friends also recognizes that not all those involved in the process of acting as a critical friend need to define the meanings of a particular data set in the same way as they can be positioned, and indeed embraced, differently in relation to their theoretical interests, research experience, power resources, and so forth. Viewed in this way, the different perspectives offered by critical friends in contrast to inter-rater reliability are positioned as a resource for challenging and developing the interpretations made by any one researcher as they construct, not find or discover through consensus, a coherent and theoretically sound argument to construct, support and defend the case they are making in relation to the data generated in a particular study. As such, despite possible agreements and disagreements, a case can be seen as defendable and the interpretation offered can be accepted as plausible. Importantly, dialogue with critical friends acknowledges that other and/or additional plausible interpretations of the data can exist that are also defendable but are not being utilized in a particular study or at that time.

Universal Criteria

The final aspect of rigor we will discuss is the notion of universal criteria, which is often drawn upon within sport and exercise psychology as an indicator of research quality

and rigorous work. Rather than simply relying on citing Lincoln and Guba (1985), in recent years the paper by Tracy (2010) has been extensively used in sport and exercise psychology to make decisions about rigor and make claims about how the quality of qualitative research was enhanced. As Burke (2016) noted when outlining qualitative research in sport and exercise, "Tracy is now on the verge of becoming the new 'benchmark' for judging all qualitative research within this sub discipline" (p. 333). However, as she also notes, there are problems with Tracy's position on criteria along with how this criteria is often applied by researchers in sport and exercise psychology.

While it needs acknowledging that Tracy (2010) speaks of criteria that "leaves space for dialogue, imagination, growth and improvisation" (p. 837), she often makes it very clear that that she advocates "eight *universal* hallmarks for high quality qualitative methods across paradigms" (p. 837; emphasis in original). It is beyond our scope to detail all of the universal criteria Tracy names. Suffice to say these are (1) worthy topic, (2) rich rigor, (3) sincerity, (4) credibility, (5) resonance, (6) significant contribution, (7) ethics, and (8) meaningful coherence. By advocating these criteria as universal markers of quality for qualitative research Tracy connects with a criteriological approach. This refers to the belief that criteria for judging qualitative research needs to be, and can be, predetermined, permanent and applied to any form of inquiry regardless of its intents and purposes (Burke, 2016; Schinke et al., 2013; Sparkes & Smith, 2009, 2014).

Taking a criteriological approach can be *strategically* useful to convince particular audiences on certain occasions (e.g. when being interviewed for a university lectureship/assistant professor by a panel exclusively made up of post-positivists) about the quality of one's research. Yet, the idea that criteria can be universally applied to all forms of qualitative research is problematic. For example, given a world of multiple, created, mind-dependent realities, and the impossibility of theory-free knowledge, criteria is not 'out there'

awaiting discovery but socially constructed. The upshot of this is that the usefulness of criteria can change and the number of criteria used in each project be modified for certain purposes. The idea of universal criteria as 'stable' markers of quality thus starts falling apart to sort out trustworthy interpretations from untrustworthy ones (see Lincoln, 2010; Smith & Deemer, 2000; Sparkes & Smith, 2009; 2014).

Applying criteria in a universal manner is also inherently problematic because it calls on a researcher to judge any piece of qualitative research, regardless of its intents and purposes, in preordained and set ways. Under such conditions, universal criteria operates in an exclusionary and punitive manner to produce a closed system of judgment that establishes and maintains a narrow band of what constitutes good research (Sparkes & Smith, 2009). Consequently, novel or different forms of research that could produce new knowledge and make a difference in society can, by definition, be excluded and/or demeaned as not worthy of attention from the outset. Much then could be lost in terms of 'additional knowledge' or producing 'impactful research'. Further, when a preordained and fixed quality appraisal 'checklist' is used, research risks becoming stagnant, insipid, and reduced to a technical exercise (Burke, 2016). There is also the risk that with the power of universal criteria to police what counts as quality research, set markers of quality become a 'strategic game' to be played by researchers to enhance publication chances. Despite the problems, some researchers might strategically use the criteria (e.g. member checking or inter-rater reliability) used in past published work within a certain journal they are targeting, and write that they used 'Tracy's criteria', to improve their chance of publishing.

Researchers might of course disagree that universal criteria is problematic (Gordon & Patterson, 2013). They might wish to continue committing to universal criteria that underpins a criteriological approach and Tracy's (2010) work. If they do continue to utilize the notion of universal criteria, they must though follow through with their commitment by adopting *all*

of the criteria that Tracy (2010) highlighted. All must be chosen since, when positioned as universal criteria, the entire eight are of equal importance as markers of quality. To make a judgment about which of these criteria should be left out, and which should be selected, would undermine universality and a criteriological approach. Yet, within sport and exercise psychology, researchers very rarely, if ever, use or demonstrate all eight criteria. What appears in publications is that researchers say they applied a select number criteria from the eight Tracy proposed as universal. They do not then follow through on their commitment to universal criteria and a criteriological approach. Given that problem, for researchers who wish to adopt universal criteria and cite Tracy in support of such use, future work should use all eight criteria. They also must take in account the numerous means, practices, and methods through which to achieve each of the eight criteria (e.g., spending enough time collecting data, thick description, self-reflexivity about subjective biases, naturalistic generalizations, and relational ethics). Further, researchers must make clear not only the approach they take (i.e. criteriological), but additionally the epistemology and ontology that underpin their work. This clarification will help reviewers and readers to judge if the research is epistemologically and ontologically cohesive from start to finish. Including the approach taken, along with the epistemology and ontology that underpins the research, is vital to include so that reviewers and readers are better placed to make fair, appropriate, and informed judgments about the quality of the research.

Finally, it should be noted that researchers in sport and exercise psychology often write in the methods section that they have applied or used ‘Tracy’s (2010) criteria’ to enhance the rigor of their work. For example, researchers often write: ‘To enhance the trustworthiness or quality of the research Tracy’s (2010) criteria was used’. Yet, Tracy did not create nor does she own the criteria presented (Burke, 2016). She synthesized the criteria highlighted in her paper from others’ work. Improper use of language and citations aside

from those in sport and exercise psychology, recognizing that Tracy neither created nor owns the eight criteria she synthesized opens up the possibility for researchers to still use, depending on the starting points, intentions and purposes of this research being judged, various markers of quality she highlighted without accepting her claims for universality. That possibility is supported when one moves away from subscribing to a criteriological approach to conceptualizing rigor through a relativist approach (Sparkes & Smith, 2009).

A relativist approach, or what is sometimes referred to as non-foundational, views criteria as a socially constructed list of characteristics (Burke, 2016; Smith & Deemer, 2000; Sparkes & Smith, 2009; Schinke et al., 2013). It means that when judging the quality of qualitative work researchers use criteria from lists that are not fixed, rigid, or predetermined before the study, but rather are open-ended; they can add to or subtract characteristics from the lists. Lists are necessarily open-ended because the criteria used can change depending upon the starting points, context and purposes of the specific piece of research being judged. That is, as various scholars argue (e.g. Levitt et al., 2016; Smith & Demmer, 2000; Sparkes & Smith, 2009) researchers might apply different criteria as they go about the practical task of judging different studies. For example, when engaging with a certain piece of work a researcher or reviewer might draw on the following list of criteria, some of which are highlighted by Tracy (2010): member reflections, critical friends, the worthiness of the topic, the significant contribution of the research, naturalistic generalizations, and what Levitt et al. (2016) described as fidelity to the subject matter (i.e. the process by which researchers develop and maintain allegiance to the phenomenon under study as it is conceived within their tradition of inquiry) and utility (i.e. the effectiveness of the research design and methods, and their synergistic relationship, in achieving study goals). Or, for different research, such as an autoethnography or creative non-fiction, they might select criteria like evocation, sincerity, aesthetic merit, expression of a reality, the meaningful contribution of

the work, incitement to action, and coherence (see Smith, McGannon, & Williams, 2015).

Accordingly, those who adopt a relativist approach do not believe that 'anything goes' as judgments always need to be made about research. However, a relativist is unwilling to mandate what one *must* do across all contexts and on all occasions prior to any piece of research being conducted. Of course, just as a researcher who commits to universal criteria needs to do, researchers who commit to a relativist approach should inform readers that relativism underpins their work and document their epistemology and ontology. It is equally vital that the execution of the study – from the data collection methods, analysis as couched within a particular methodology, ethics, and write-up of results, aligns with said criterion, epistemology, and ontology (Sparkes & Smith, 2009). Recent examples of researchers in sport and exercise research using a relativist approach can be found in Papathomas and Lavallee (2014), Hill and Hemmings (2015), Smith, Bundon and Best (2016), Coyle, Gorczynski, and Gibson (2017), and Coppola, Dimler, Letendre, and McHugh (2017).

Conclusions

Qualitative research, once on the fringes, now plays a central part in advancing sport and exercise psychology knowledge and producing impactful research (Kay, 2016). Having provided a strong rationale for the need for researchers to engage in contemporary methodological thinking, the present paper adds to the qualitative literature by offering a detailed review of three widely used ways to demonstrate rigor within sport and exercise psychology. It was argued qualitative work in the field often rests on problematic methods and outdated ideas when it comes to developing rigorous qualitative research. It is vital then that we avoid perpetuating the problems with how rigor has been commonly sought and evaluated in the past. If then our field is to produce high quality research, it is necessary that scholars, including journal editors and reviewers, change how rigor is developed and qualitative work is judged. It is also important that university students at all levels are taught

by experienced qualitative researchers and trained in epistemology, ontology, methodology and up-to-date methods (Knight, 2016; McGannon & Schweinbenz, 2011; Terkildsen & Petersen, 2015).

Grounded in the arguments detailed and the contemporary literature, plus inspired by Dunnette's (1966) critique of fads, fashion and folderol in psychology research, the following summaries and recommendations for doing qualitative research in sport and exercise are offered:

1. Member checking does not ensure that the results of qualitative research are valid or trustworthy. Because theory free knowledge cannot be achieved, the method is unable to access an independent social reality in order to demonstrate that the results correspond to the reality and the truth has been objectively found. Member checking has no evidence based to support it as a verification method. There are also many practical problems that cannot be overcome, which in turn, reinforces the point that the method is unable to verify or demonstrate the trustworthiness of results. Member checking is therefore an ineffective marker to judge the rigor or quality of qualitative research. Given this, researchers may move to member reflections whereby findings are shared with participants, but with a different logic of justification, intention and outcome sought than as in member checking. For example, in member reflections objectively accessing the truth or getting at the reality independent of us via methods is not viewed as a problem of qualitative research; it is simply part of our human finitude that must be accepted as part of the research. Member reflections might be useful to produce rigorous qualitative research but not then in the sense of verifying research by getting at an independent truth. Reflections offered by participants are instead reframed as a way to help create a meticulous, robust, and intellectually enriched understanding through generating additional insights and dialogue. In certain circumstances member reflections may also be

used to promote ethical practice.

2. Inter-rater reliability as traditionally used (Lincoln & Guba, 1985) in research via the practice of intercoder reliability and intercoder agreement is ineffective for ensuring that results are reliable. Like with member checks, researchers should therefore give up using that method as a way to ensure rigor. This recommendation is made because theory-free knowledge is not possible in the process of engaging in intercoder reliability and intercoder agreement. Also there are various problems with inter-rater reliability, such as unitization and there being no numerically agreed level of what counts as a ‘good agreement’ between coders to sort out the more rigorous from the less rigorous research. If researchers believe reliability is still a concern for their work then it is recommended that new inter-rater reliability guidelines be tested and adopted (e.g., MacPhail et al., 2015). If researchers conclude that reliability is not for good reasons a relevant concern for evaluating qualitative research, they need to look elsewhere for markers of rigor to ensure the research is of high quality. As part of a list of characterizing traits for quality and rigor, this might include adopting critical friends, which is viewed as an opportunity for dialogue and the reflexive acknowledgement of multiple truths, perspectives and results in the research process.
3. If a researcher subscribes to the notion of universal criteria and a criteriological approach as a way of ensuring rigor, and follows Tracy (2010), they must demonstrate the use all the universal criteria as well as the various markers or methods to achieve each marker of quality that she highlighted. If one accepts the problems of adopting universal criteria and a criteriological approach then a relativist approach, which advocates using criteria in a list-like like manner, is a viable option. No matter what approach is used researchers must be clear which one is adopted and the implications for the research process that follow (i.e., from data collection methods, methodology, analysis and results presentation). In so

doing, researchers must also make clear the epistemology and ontology that underpins their work as well as the criteria used, whether that is universal or criteria/list-like.

As an invitation to developing qualitative research, it is hoped that this paper enables researchers - including editors and reviewers - to understand the various problems with certain methods and move to other possibilities to enhance the quality of qualitative research within sport and exercise psychology. Of course, in one paper we could never cover the numerous debates that are occurring on qualitative research and how we can develop our understandings further within and across all of these debates and discussions. For example, interested readers may refer to important points recently been raised about how qualitative research can be generalizable not in a statistical sense but in other ways (Sparkes & Smith, 2014), doing a thematic analysis (Braun, Clarke, & Weate, 2016), problematic assumptions in mixed methods research (McGannon & Schweinbenz, 2011; Gibson, 2016; Sparkes, 2015), the cherry picking of grounded theory techniques along with a failure to demonstrate the commonalities and variants of grounded theory in work (Holt, 2016; Weed, 2017), post-qualitative research (Fullagar, 2017; Giardina, 2017), and the value of exceptions within qualitative data (Phoenix & Orr, 2017). We hope that qualitative researchers in the future will not only reflexively and critically engage with such debates occurring, but also contribute to these in order to expand upon the numerous dialogues on qualitative research within the field.

Acknowledgements

Thank you to the anonymous reviewers for their insightful comments that enhanced the paper.

References

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member checking: A tool to enhance trustworthiness or merely a nod to validation? *Qualitative Health Research*, 26, 1802-1811.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. London: Sage.

Braun, V., Clarke, V., & Weate, P. (2016). Using thematic analysis in sport and exercise research. In B. Smith, & A. Sparkes (Eds.*), Routledge handbook of qualitative research methods in sport and exercise* (pp. 191-205). London: Routledge.

Burke, S. (2016). Rethinking 'validity' and 'trustworthiness' in qualitative inquiry: How might we judge the quality of qualitative research in sport and exercise sciences? In B. Smith, & A. C. Sparkes (Eds.), *Routledge handbook of qualitative research in sport and exercise* (pp. 330-339). London: Routledge.

Campbell, J.L. Quincy, C., Osserman, J., & Pedersen, O. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods and Research*, 42, 294-320.

Communication Studies 298 (2000). A night with the narcissist and the nasty boys: Interpreting the World Wrestling Federation. *Qualitative Inquiry*, 6, 526-545.

Cook, K.E. (2011). Reliability assessments in qualitative health promotion research. *Health Promotion International*, 27, 90-101.

Cosh, S., LeCouteur, A., Crabb, S., & Kettler, L. (2013). Career transitions and identity: A discursive psychological approach to exploring athlete identity in retirement and the transition back into elite sport. *Qualitative Research in Sport, Exercise and Health*, 5, 21-42.

Coppola, A.H., Dimler, A.J., Letendre, T.S., & McHugh, T-L. F. (2017). 'We are given a

body to walk this earth': The body pride experiences of young Aboriginal men and women. *Qualitative Research in Sport, Exercise and Health*, 9, 4-17.

Coyle, M., Gorczynski,P., & Gibson, K. (2017). "You have to be mental to jump off a board any way": Elite divers' conceptualizations and perceptions of mental health. *Psychology of Sport and Exercise*, 29, 10-18.

Cowan, D., & Taylor, I.M. (2016). ''I'm proud of what I achieved; I'm also ashamed of what I done': A soccer coach's tale of sport, status, and criminal behavior. *Qualitative Research in Sport, Exercise and Health*, 8, 505-518.

Culver, D., Gilbert, W., & Sparkes, A.C. (2012). Qualitative research in sport psychology journals: The next decade 2000-2009 and beyond. *The Sport Psychologist*, 26, 261-281.

Denzin, N.K. (2017). Critical qualitative inquiry. *Qualitative Inquiry*, 23, 8–16.

Dunnette,, M.D. (1966). Fads, fashions and folderol in psychology. *American Psychologist, 21*, 343-352.

Eakin, J.M. (2016). Educating critical qualitative health researchers in the land of the randomized controlled trial. *Qualitative Inquiry*, 22, 107-118.

Estroff, S. E. (1995). Whose story is it anyway? Authority, voice and responsibility in narratives of chronic illness. In S. K. Toombs, D. Barnard, & R. A. Carson (Eds.), *Chronic illness: From experience to policy* (pp. 77–104). Bloomington: Indiana University Press.

Fullagar, S. (2017). Post-qualitative inquiry and the new materialist turn: Implications for sport, health and physical culture research. *Qualitative Research in Sport, Exercise and Health*, 9, 247-257.

Gergen, K. (2009). *Relational being*. Oxford: Oxford University Press.

Giardina, M.D. (2017). (Post?)qualitative inquiry in sport, exercise, and health: Notes on a

methodologically contested present. *Qualitative Research in Sport, Exercise and Health*, 9, 258-270.

Gibson, K. (2016). Mixed methods research in sport and exercise. In Smith, B. & Sparkes, A. C. (2016) (Eds). *Routledge handbook of qualitative research methods in sport and exercise* (pp. 382-396). London: Routledge.

Gordon, J., & Patterson, J. (2013). Response to Tracy's under the "Big Tent": Establishing universal criteria for evaluating qualitative research. *Qualitative Inquiry*, 19, 689-695.

Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. London: Sage.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 191-216). Thousand Oaks, CA: Sage.

Hill, D., & Hemmings, B. (2015). A phenomenological exploration of coping responses associated with choking in sport. *Qualitative Research in Sport, Exercise and Health*, 7, 521-538.

Holt, N. L. (2016). Doing grounded theory in sport and exercise. In B. Smith & A. C. Sparkes (Eds.), *Routledge handbook of qualitative research in sport and exercise* (pp. 24-36). London: Routledge.

International Conference for Qualitative Sport and Exercise (2016, December 17th). Retrieved from twitter @QRSE2018

Jones, I., Brown, L., & Holloway, I. (2012). *Qualitative research in sport and physical activity*. London: Sage.

Kay, T. (2016). Knowledge, not numbers: Qualitative research and impact in sport, exercise and health. In Smith, B. & Sparkes, A. C. (2016) (Eds). *Routledge handbook of qualitative research methods in sport and exercise* (pp. 424-437). London: Routledge.

Knight, C.J. (2016). Teaching qualitative research. In Smith, B. & Sparkes, A. C. (2016) (Eds). *Routledge handbook of qualitative research methods in sport and exercise* (pp. 409-423). London: Routledge.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. London: Sage.

Levitt, H. M., Motulsky, S. L., Wertz, F. J., Morrow, S. L., & Ponterotto, J. G. (2016). Recommendations for designing and reviewing qualitative research in psychology: Promoting methodological integrity. *Qualitative Psychology*. Advance online publication. http:// dx.doi.org/10.1037/qup0000082

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lincoln, Y.S. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry*, 1(3), 275–289.

Lincoln, Y.S. (2010). 'What a long strange trip it's been…': Twenty-five years of qualitative and new paradigm research. *Qualitative Inquiry, 16,* 3–9.

MacPhail, C., Khoza, N., Abler, L and Ranganathan, M (2016). Process guidelines for establishing Intercoder Reliability in qualitative studies. *Qualitative Research*, 16,198–212.

Madill, A., Jordan, A., & Shirley, C. (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *British Journal of Psychology*, 91, 1-20.

McGannon, K.R. & Schweinbenz, A. (2011). Traversing the qualitative-quantitative divide using mixed methods: Some reflections and reconciliations for sport and exercise psychology. *Qualitative Research in Sport, Exercise and Health, 3,* 367-381.

McGannon, K.R., & Smith, B. (2015). Centralizing culture in cultural sport psychology research: The potential of narrative inquiry and discursive psychology. *Psychology*

of Sport and Exercise, 17, 79-87.

McGannon, K. R., & Spence, J. C. (2010). Speaking of the self and physical activity participation: What discursive psychology can tell us about an old problem. *Qualitative Research in Sport and Exercise, 2,* 17-38.

Morse, J. M. (1997). Perfectly healthy, but dead: The myth of inter-rater reliability. *Qualitative Health Research*, 7, 445–447.

Morse, J. M. (2016). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research*, 25, 1212-1222.

Palmer, C. (2016). Ethics in sport and exercise research: From research ethics committees to ethics in the field. In Smith, B. & Sparkes, A. C. (2016) (Eds). *Routledge handbook of qualitative research methods in sport and exercise* (pp. 316-329). London: Routledge.

Papathomas, A., & Lavallee, D. (2014). Self-starvation and the performance narrative in competitive sport. *Psychology of Sport and Exercise*, 15, 688-695.

Phoenix, C., & Orr, N. (2017). Analysing *exceptions* within qualitative data: Promoting analytical diversity to advance knowledge of ageing and physical activity. *Qualitative Research in Sport, Exercise and Health*. Advance online publication. Doi: 10.1080/2159676X.2017.1282539

Pitney, W., & Parker, J. (2009). *Qualitative research in physical activity and the health professions*. Champaign, IL: Human Kinetics.

Saldaña, J. (2013). *Fundamentals of qualitative research*. Oxford: Oxford University Press.

Schinke, R.J., McGannon, K.R. & Smith, B. (2013). Expanding the sport and physical activity research landscape through community scholarship. *Qualitative Research in Sport, Exercise and Health, 5,* 287-290.

Smith, B., Bundon, A., & Best, M. (2016). Disability sport and activist identities: A qualitative study of narratives of activism among elite athletes' with impairment.

Psychology of Sport and Exercise, 26, 139-148.

Smith, B., McGannon, K.R., & Williams, T. (2015). Ethnographic creative non-fiction: Exploring the what's, why's and how's. In L. Purdy & G. Molner (Eds.). *Ethnographies in sport and exercise* (pp. 59-73). London: Routledge.

Smith, B., & Sparkes, A. C. (2016a) (Eds). *Routledge handbook of qualitative research methods in sport and exercise*. London: Routledge.

Smith, B. & Sparkes, A. C. (2016b). Introduction: An invitation to qualitative research. In Smith, B. & Sparkes, A. C. (2016) (Eds). *Routledge handbook of qualitative research methods in sport and exercise* (pp. 1-7). London: Routledge.

Smith, B., & Sparkes, A. C. (2016c). Interviews: Qualitative interviewing in the sport and exercise sciences. In B. Smith & A. C. Sparkes (Eds.), *Routledge Handbook of Qualitative Research Methods in Sport and Exercise* (pp. 103-123). London: Routledge.

Smith, J., & Deemer, D. (2000). The problem of criteria in the age of relativism. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 877–896). London: Sage.

Smith, J., & Hodkinson, P. (2005). Relativism, criteria and politics. In N. Denzin, & Y. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed.). (pp. 915–932) London: Sage.

Smith, J., & Hodkinson, P. (2009). Challenging neorealism: A response to Hammersley. *Qualitative Inquiry*, 15, 30–39.

Sparkes, A.C. (2015). Developing mixed methods research in sport and exercise psychology: Critical reflections on five points of controversy. *Psychology of Sport and Exercise*, 16, 49-59.

Sparkes, A.C., & Smith, B. (2009). Judging the quality of qualitative inquiry: Criteriology

and relativism in action. *Psychology of Sport and Exercise, 10*, 491-497.

Sparkes, A.C., & Smith B. (2014). *Qualitative research methods in sport, exercise and Health: From process to product*. London: Routledge.

Terkildsen, T., & Petersen, S. (2015). The future of qualitative research in psychology—A students' perspective. *Integrative Psychological and Behavioral Science*, 49, 202–206.

The European Federation of Sport Psychology (2016, December 17th). Retrieved from http://www.fepsac.com.

Thomas, D.R. (2017). Feedback from research participants: Are member checks useful in qualitative research? *Qualitative Research in Psychology*, 14, 23-41.

Tracy, S.J. (2010). Qualitative quality: Eight "big tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*, 837-851.

Weed, M. (2017). Capturing the essence of grounded theory: The importance of understanding commonalities and variants. *Qualitative Research in Sport, Exercise and Health, 9*, 149-156.

Wolcott, H. (1994). *Transforming qualitative data*. London: Sage.

Wolcott, H. (1995). *The art of fieldwork*. London: AltaMira Press.

Young, K., & Atkinson, M. (Eds.) (2012). *Qualitative research on sport and physical culture*. Bingley, UK: Emerald Group Publishing Ltd.

---

[1] A common misunderstanding attached to believing in epistemological constructionism and ontological relativism is that a researcher denies that there is a physical world out there independent of us. This is not the case; a *physical* world is accepted (Smith & Hodkinson, 2009). But, while it is believed that there are physical beings out there moving around in time and space and uttering what people call words, it is argued that the

interpretations/descriptions we offer of these movements and utterances are not out there in the sense of being independent of our interests and purposes.

[2] If a researcher does believe that recall bias or member distortion is a problem when conducting interviews then to write this as a limitation in a paper that uses interviews reflects a poor decision on their behalf. From the start, they should not have chosen interviews but rather, for example, naturalistic data (i.e. data that is generated without the influence of the researcher) (Smith & Sparkes, 2016c). The ‘problem’ and ‘limitations’ then lies with their methodological decision making.