

Challenging the New Orthodoxy: A Critique of SPLISS and Variable-Oriented Approaches  
to Comparing Sporting Nations

Abstract

**Research Question:** In recent years the comparative sport policy field has become dominated by the ‘SPLISS’ approach developed by De Bosscher and colleagues. While the SPLISS approach has developed important insights into the statistical relationship between key groups of independent variables and indicators of elite sport policy success, nevertheless its attempts to identify and explain both statistical association and causal relationships have significant limitations.

The paper thus seeks to address the question of the nature of such strengths and limitations and their implications for theory, policy and practice.

**Methods:** As a review paper it develops a critical evaluation of claims made for the SPLISS approach to variable oriented comparative policy analysis.

**Results:** The paper identifies and focuses on the implications of six key problems for the SPLISS approach, namely: philosophical assumptions and causal variables; the black box problem; internal validity issues; non-equivalence and reliability; the neglect of agency; and misconceptions in the use of mixed methods.

**Implications:** The paper’s findings represent a challenge to the hegemony of this variable-oriented approach and they argue not for replacement or rejection of such an approach, but for recognition of its limitations, and of the opportunities for complementing it with case-driven, qualitative analysis generating causal accounts of policy outcomes.

*Keywords:* sport policy, comparative, methodology; elite sport, sporting nations, SPLISS

## **Challenging the New Orthodoxy: A Critique of SPLISS and Variable-Oriented Approaches to Comparing Sporting Nations**

Approaches to understanding the factors which lead to international sporting success, and in particular Olympic medal success, have evolved over time, but have intensified in the last two decades (Andersen & Roglan, 2012; Bergsgard, Houlihan, Mangset, Nødland, & Rommetveldt, 2007; De Bosscher, Bingham, Shibli, Van Bottenburg, & De Knop, 2008; De Bosscher, De Knop, Van Bottenburg, & Shibli, 2006; De Bosscher, Shibli, Westerbeek, & Van Bottenburg, 2015; Digel, Barra, Bosel, Fahrner, Kohl, & Utz, 2002; Digel, Burk, & Fahrner, 2006). These studies might be categorised as: critical description of the development of elite sport systems; input-throughput-output studies which calibrate statistically the relationship between investment and sporting success; and meso-level theory which seeks to understand the policy processes surrounding elite sport development.

Early comparative sports studies focused on the systems and practices adopted by the former Communist bloc, and in particular the Soviet Union and the German Democratic Republic to produce elite sport success. Riordan and others (Riordan, 1999; Riordan & Jones, 1999), constitute critical description of the philosophy and operationalisation of support for elite sport development systems highlighting particular issues such as talent identification and specialization at an early age, specialist sport schools, sports science support, and doping strategies.

Henry, Amara, Al-Tauqi, and Lee (2005) introduce a typology of four methodological approaches to comparative analysis: Statistical comparison ('Seeking Statistical Similarities'), qualitative description ('Describing Differences'), embedding accounts of national policy in the context of theoretical accounts of change in the transnational environment ('Theorising the Transnational'), and the discursive construction of policy

systems ('Defining Discourse'). Each approach relies on different ontological and epistemological assumptions and thus engages in a comparative analysis of a significantly different order, each with their own strengths and weaknesses. Chalip et al.'s (1996) analysis constitutes an example of 'Describing Differences', and can be contrasted with Digel et al.'s (2002) who effectively seeks to establish comparison on the basis of common data measures for a single sport (athletics), and constitutes an example of 'Seeking Statistical Similarities'. There is also a range of studies examining the statistical relationship between macro factors such as GDP, population size, political system, degree of urbanisation etc. which are not necessarily directly related to, even if they are statistically associated with, elite sport success (De Bosscher, 2007).

In terms of studies that reflect statistical relationships between characteristics of national context and elite sport performance, De Bosscher and colleagues confirm earlier claims of the significance of GDP per capita and population size in explaining variance in Olympic medal winning performance by employing stepwise regression analysis. Their analysis indicates that 52.4% of elite sport success can be attributed to population size, national wealth and whether the nations' political system was part of the former communist bloc (De Bosscher, Heyndels, De Knop, van Bottenburg, & Shibli, 2008, p. 225).

The authors go on to argue that the residuals for each country (which represent the unexplained variance in the dependent variable of the regression analysis) incorporate the impact of sports policies aimed at securing medal success – though it is not possible to say what proportion of the unexplained variance, is attributable to such factors, and indeed 52.4% of variance explained by these macro factors may already incorporate an element of the variance explained by policy factors. For example, a variable such as GDP per capita may be related to elite sports policy variables such as the level of financial investment in elite sport (richer countries are likely to have greater 'disposable income' to invest in sport) but it is not

the fact of national wealth which might explain performance, but rather what proportion of national resources are allocated to the elite sport system and how such investment is used, which will provide rational explanation of why performance varies between nation states.

Digel et al.'s (2002) study is, in a sense, a forerunner of what the SPLISS studies have taken forward in a more comprehensive fashion utilising the input-throughput-output approach. Digel (2002) adopted a mixed-method approach to compare elite level systems of track and field athletics in eight countries. Their account incorporates a relatively transparent summary of the difficulties conducting comparative analysis. These problems included the German-centric assumptions of the team of researchers which limited both the nature of the questions to be raised and the terms in which such questions are framed and understood, as well as the lack of comparability of statistical data (Digel et al., 2002, p. 78).

The work of Green and Oakley (2001) was also a precursor to the SPLISS studies, in that the authors sought to establish key success factors associated with international sporting success, and in particular the transfer of policy learning in the production of elite sporting success from the former eastern bloc states. However, this analysis, unlike the SPLISS studies, represents a qualitative case-based analysis of broadly comparable countries, rather than an analysis of variables statistically associated with elite sport performance outcomes.

Green and Houlihan (Green, 2004; Green & Houlihan, 2004, 2005; Houlihan, 2013; Houlihan & Green, 2008) develop further explanations of the adoption of particular elite sport policy approaches by drawing upon meso-level theories of policy learning and transfer, including application of the Advocacy Coalition Framework (ACF) (Sabatier & Jenkins-Smith, 1993), path dependency theory (Kay, 2005; Peters, Pierre, & King, 2005), mimetic isomorphism (DiMaggio & Powell, 1983), and regime theory (Houlihan, 2009). Similarly, Andersen and Ronglan (2012) and Bergsgard et al. (2007) adopt institutional frameworks to examine the similarity and convergence of elite sport policies in Nordic and western nations

respectively. Nevertheless, although such approaches provide insights into the policy process, they do not generally address policy outcomes or causal relationships between inputs, throughputs, outputs, and outcomes in the manner explicitly adopted by the SPLISS studies.

What is unique about the SPLISS studies is their attempt to generate evidence of causal relations between particular policies, or groups of policies, and elite sport outcomes. The SPLISS studies adopt the rhetoric and the conceptual architecture of input-throughput-output-outcome analysis which has evolved over time. This evolution is evidenced in the two major waves of studies (SPLISS 1.0, and SPLISS 2.0) initiated by the SPLISS team. SPLISS 1.0 constituted a six country study involving Belgium (separated into data for Flanders and Wallonia), Canada, Italy, the Netherlands, Norway, and the United Kingdom (De Bosscher, et al., 2008; De Bosscher, De Knop, & van Bottenburg, 2009; De Bosscher, De Knop, van Bottenburg, Shibli, & Bingham, 2009; De Bosscher, Shibli, van Bottenburg, De Knop, & Truyens, 2010). SPLISS 2.0 (De Bosscher et al., 2015) is based on a wider study of 15 nations, composed of three of the SPLISS 1.0 nations, namely Belgium (Wallonia and Flanders), Canada, the Netherlands, and 12 others (Denmark, Estonia, Finland, France, Northern Ireland (UK), Portugal, Spain, Switzerland, South Korea, Japan, Australia, and Brazil).

The broad strategy of SPLISS 1.0 is much more explicitly positivistic than the 2.0 version which sought to expand its approach through adopting what its authors describe as a mixed methods approach, though we contest this claim below. SPLISS 1.0 set out to find the unique route to sporting success, ignoring the possibility of equifinality, that is of reaching the same endpoint in more than one way.

However, although De Bosscher and colleagues recognise that there is no single configuration which uniquely generates elite sport success (see De Bosscher et al., 2015, p. 5), SPLISS 2.0 still pursues the goal of providing a one size (or one sample) approach to the

problem, suggesting that better answers will be found through a larger sample of countries, more highly developed instruments, and a larger number of dedicated researchers (*ibid*, p. 15). This approach misses the critical point that national elite sports policy systems are not all drawn from the same population. Some will be from the top echelon of medal winning nations whereby the number of medals won, or market share, might be a broadly appropriate dependent variable or measure of elite sporting success. Others will be drawn from the group of nations which, while not necessarily competing for many medals, nevertheless specialize in particular sports (e.g. Kenya/Ethiopia in distance running), while others still may have little or no record of medal-winning, and other measures of elite success would have to be adopted for these nations (e.g. performance at regional level competitions).

Notwithstanding the advances which have been made in terms of sophistication of analysis, there are a number of underlying limitations with the SPLISS approach that remain. These problems are probably best described as limitations, in the sense that they do not invalidate the analysis premised upon the SPLISS approach, but rather they highlight the limits of the claims which might be made on the basis of the SPLISS approach. The purpose of this paper, therefore, is to offer a qualified critique of variable-orientated approaches to comparative sport policy analysis in general and to identify a number of limitations of the SPLISS approach specifically. We outline six interrelated limitations of the SPLISS approach below.

## **Six Key Limitations of the SPLISS Approach**

### **Problem 1: Philosophical Assumptions and the Identification of Causal Variables**

The key goal of the SPLISS studies is to identify the critical success factors explaining elite sport success. This is operationalized by establishing the statistical relationship between a dependent variable usually expressed as numbers of Olympic medals won (De Bosscher et

al., 2015, p. 87), medal market share (Shibli, & Bingham, with Henry, 2007), losing Olympic finalists / diplomas (De Bosscher et al., 2015, p. 101), and groups of independent variables described as ‘pillars’. Thus, the approach for SPLISS 1.0 and 2.0 has been to seek to establish the ‘facts’ of the causes of elite sport outcomes by reference to statistical associations of aggregated scores of Critical Success Factors (CSF’s) for each pillar (based on ‘facts’ and stakeholder perceptions). According to De Bosscher and colleagues, “the SPLISS study was driven by the need to compare a large amount of data to gain insights into the analytical relationships between policy (input-throughput) and success parameters (outputs)” (De Bosscher et al., 2015, p. 3).

The authors claim that their approach is more than one which seeks to establish statistical association between variables, and the implication that knowing what inputs are made and are associated with particular outputs does not allow us to claim that we can identify causes of particular outcomes. Knowledge that particular inputs are *associated with* given outcomes is no substitute for understanding / explaining *how* inputs are used to achieve or *cause* particular outcomes. Theory-driven approaches of success in achieving policy outcomes offer greater heuristic potential for explanation of such outcomes. Realist evaluation (e.g. Pawson, 2006, 2013; Pawson & Tilley, 2004), for example, focuses on causal mechanisms and their relationships with the local social, cultural, political, economic and organisational context, using theories of social change and / or logic models (Cooksy, Gill, & Kelly, 2001; Jordan, 2010) to inform ways of testing of theory-informed causal accounts of such change.<sup>1</sup> The point that explanation of change cannot be achieved by statistical

---

<sup>1</sup> While Pawson’s position is realist he rejects the critical theory stance implied in Bhaskar’s critical realism ARCHER, M., BHASKAR, R., COLLIER, A., LAWSON, T. & NORRIE, A. (eds.) 1998. *Critical Realism: essential readings*, Abingdon, Oxon: Routledge., opting instead for what he terms a middle range realism PAWSON, R. 2006. *Evidence-based policy : a realist perspective*, London, SAGE, PAWSON, R. 2009. Middle Range Realism. *European Journal of Sociology / Archives Européennes de Sociologie*, 41, 283-325..

association alone is acknowledged by the advocates of SPLISS (see De Bosscher, 2006 and De Bosscher et al., 2010) but not successfully addressed.

Both SPLISS 1.0 and 2.0 seek to elicit qualitative data which can reveal aspects of the internal mechanisms at play within the elite sport policy system. However, the nature of the data collected, and the manner in which data were collected, analysed, and employed (and the lack of detail in terms of protocols for how this has been done) make this element of the evidence appear to be far from systematic and rigorous, and thus their contribution to explanation is unconvincing.

The development of causal accounts which explain the outcomes of medal success would benefit from a more rigorous and developed causal framework. Such frameworks have been advanced in the policy analysis field, in for example the realist approach of Pawson and Tilley (Pawson, 2006; Pawson & Tilley, 1997). There are a number of benefits to this approach. The realist approach characterizes social policy contexts as open systems, explanations of which are required to place a greater emphasis on contextual factors in explaining outcomes. The approach which is summarized in Pawson and Tilley's 'formula'  $C+M=O$ : Context (C) plus causal Mechanism (M) explain Outcomes (O), implies that the Critical Success factors which the SPLISS approach rests upon, may well operate differently in different contexts. Of course SPLISS acknowledges the importance of economic and political contexts (the importance for example of the former communist, state-controlled athlete production systems; or of the neo-liberal sports economies such as that of the United States; or state intervention along the lines of social welfare economies, such as those of the Scandinavian welfare states). However, while the SPLISS team acknowledges the significance of different configurations of variables in different contexts it nevertheless underplays in its own explanations of cultural differences (whether generic, or sporting). In short, the SPLISS approach is seeking the answer to 'what works' in terms of producing elite



sport success, whereas realist analysis addresses the question of ‘what works? for whom? under which circumstances?’, developing a more nuanced account in contexts which are diverse and changing, in other words which constitute open systems.

## **Problem 2: The ‘Black Box’ Problem**

Related to the underdevelopment of causal accounts of policy change is the ‘Black Box’ problem, where the correlation between measurable inputs of resources, and outputs in terms of results may be measured but the policy processes that are responsible for turning inputs into outputs are not amenable to observation or measurement and thus are hidden from view (hence the allusion to the policy implementation system as a ‘black box’). The SPLISS authors are obviously aware of the ‘Black Box’ problem but claim to have resolved this (De Bosscher et al., 2008, p. 35).

The authors characterise their explanation as identifying inputs in pillar 1, that is largely the input of financial resources, and the use of this funding for facilitating the development of the remaining pillars which are conceptualized as throughputs. However, treating variables such as the number of coaches qualified and employed at various levels (Pillar 7) is hardly a throughput. It is not the number of coaches employed within (or input into) the system which brings about success, but rather it is how such coaches are utilized or employ their skills which explains the success of the system. Similarly, it is not the input of money into the sports science support system which explains success, but how that support system is organized to leverage success. The input is sports science knowledge and resources, the throughput is the way in which sports science is employed, the nature of the causal mechanisms identified and employed by those delivering the programme.

This issue is bound up with the nature of qualitative data in the specification of throughputs. If we are to describe or define throughputs, we are implicitly bound to the process of specifying (or at the very least speculating upon) causes of outcomes. However,

when we dig down into the ways in which qualitative data are obtained and analysed (as we note under our discussion of the misuse of the term mixed methods in discussing our sixth ‘problem’ below), there are significant limitations to the usefulness of this information.

Explaining what is happening within the policy system (what is going on within the ‘black box’) in terms of factors critical to the success of the policy system implies more than simply identifying the statistical association, i.e. the constant conjunction of input with output, or even of throughput resources (which we conceive of as inputs) and outputs. Rather it involves detailed qualitative accounts of how particular actions bring about particular outcomes, requiring detailed *qualitative analysis* of the qualitative data obtained.

### **Problem 3: Internal Validity**

Internal validity relates to the issue of whether outcomes observed in a study / analysis are due to changes in an independent variable or variables, and not to some other, external, factor(s). In other words, internal validity implies a causal relationship between the independent and dependent variable. Internal validity is related to the extent to which an operational measure captures the reality of the concept to be measured, and is said to exist if the observed variation of the independent variable(s) affects the value of the dependent variable (e.g. medals or market share).

The dependent variable for the SPLISS studies has tended to be operationalised in terms such as Olympic medals won, or ‘market share’ of medals achieved. Given that by 2018 after the Pyeongchang Games, 34% of National Olympic Committees had never won a medal in either the Winter or Summer Olympics, of any sort, and 78% had never won a medal in the Winter Games, it is clear that as a measure of ‘success’ in elite sport terms, this is a relatively blunt instrument for a large proportion of nation states. Thus, a more useful approach might be to cluster nations into appropriate performance brackets and to evaluate nations by comparison with other nations with similar sporting profiles (e.g. using dependent

variables of performance in continental or regional games and / or using resource capacities such as GDP to group cases together, and thus ‘comparing apples with apples’), or to identify success in terms of improvement of performance by self-benchmarking of nations over time. A further complication relates to the treating of medals for all sports as being of equivalent value which may not reflect national priorities, and thus may not be a culturally appropriate, or valid measure of success. Nations may also adopt more specific priorities such as placing a value on achieving equity goals (e.g. having a relatively even spread of medals across both genders in Olympic and / or Paralympic competition).

Perhaps the dominant view in relation to comparative analysis in the social sciences, is that it should be based on analysis at the level of variables. However, as Della Porta (2008: 198) argues,

... much research – especially in political science, but also in some branches of sociology – is case-oriented: that is, it aims at rich descriptions of a few instances of a certain phenomenon. ... Variable-oriented studies mainly aim at establishing generalized relationships between variables, while case-oriented research seeks to understand complex units.

As Della Porta suggests, both approaches are legitimate, but in the case of the SPLISS analysis of individual, or groups of, variables, case level analysis is neglected.

An inherent danger of the variable-oriented approach is that whilst it may be methodologically convenient to apply a ‘one-size-fits-all’ nine-pillar framework across all sporting nations, it assumes similarities of cases, overlooking (or ignoring) fundamental differences that may exist within sporting systems. Thus, the deductive application of a pre-determined framework, ignoring the development of inductive insights into the differences between cases, has implications for sampling equivalence (see limitation 4 below).

Furthermore, the assumption often made by variable-orientated researchers seeking large-N comparisons is that the greater the number of variables employed in the analysis, the ‘closer’ or ‘better’ a reflection of reality the data are assumed to provide (Landman & Carvalho, 2017). The SPLISS account has sought to increase its size and scale. In SPLISS 2.0, for example, the authors highlight that “the project involved 15 nations and responses from 3000 elite athletes, over 1300 coaches and 241 performance directors...in collaboration between 58 researchers and 33 policymakers’ (De Bosscher et al., 2015, pp. 21-22) and that “the elite sport climate survey, [was] completed by 3142 athletes, 1376 coaches and 243 performance directors” (De Bosscher et al, 2015, p. 12). The implication here is that more data implies greater explanatory power.

Whilst the SPLISS studies should be acknowledged as one of the largest empirical studies of its kind within the sport management domain, the approach adopted of expanding the numbers of nations in the sample implicitly assumes that increasing the number of nations involved in the study will enhance (confidence in) its findings. As we have already indicated the validity of medals as a measure of elite sporting success is weak because countries do not all belong to the same population as sporting nations, with, in many cases, nations not winning any medals at all. Thus different sub-populations of nations require different measures of sporting success if these are to be valid. Qualitative analysis of what constitutes valid measures of sporting success for individual or groups of nations (in particular those which do not win Olympic medals) is likely to produce more telling insights than simply increasing the numbers of variables employed in developing nomothetic models.

With regard to the creation of the nine-pillar framework itself, the SPLISS team claims this to be a conceptual model of the sports policy factors leading to international sporting success that was derived from a systematic review of existing literature (though this would seem to relate to a literature review which was conducted rigorously, rather than to the

conducting of a Systematic Review: Petticrew and Roberts, 2005). An important fundamental question remains as to whether the nine-pillar depiction derived by De Bosscher and colleagues (De Bosscher et al., 2006; De Bosscher et al., 2009) is indeed a framework or model. These words are often used inter-changeably by the research consortium. Such arguments may be viewed as semantic but are important when establishing whether the SPLISS account can claim a causal relationship between its pillars (independent variables) and the outcome variable (e.g. medals success) or between the pillars themselves. This issue is central to understanding the nature and purpose of the SPLISS account and whether it can claim to be more than an elaborate descriptive account or an attempt to classify sport policy factors. In other words, they are important in understanding whether SPLISS can claim to be predictive.

#### **Problem 4: Non-equivalence and Reliability**

Equivalence in essence refers to how we know that we are studying the same phenomena in different contexts; in other words, how we know that the concepts and instruments utilized measure the same things in different contexts (Dowling, Brown, Legg, & Grix, 2018). The issue of (non-)equivalence is a methodological challenge for all comparative researchers and it is important to recognize that equivalence issues are pervasive at all levels of comparative analysis (Øyen, 2004; Ebbinghaus, 2005; Dowling et al., 2018; Jowell, 1998). However, we would argue that it is particularly problematic for large-scale, variable-orientated comparative studies like SPLISS.

De Bosscher and colleagues assume construct, functional and sampling equivalence. Construct equivalence refers to ensuring the instruments employed measure the same variables across nations. This is particularly challenging for sport policy scholars and as taken-for-granted concepts (e.g. athlete, coach, elite, participation, sport) in one country or

language, are not necessarily equivalent in another as Digel and colleagues (2002) discovered.

Another fundamental issue surrounding equivalence is the compatibility and comparability of the secondary data used within the SPLISS studies. The comparative methodology literature refers to this as functional equivalence i.e. whether the data collected is standardized and contextualized to allow for meaningful comparisons (e.g., Øyen, 2004; Ebbinghaus, 2005; Landman & Carvalho, 2017). Comparative studies even of financial data on sport, within relatively similar governmental and economic systems have manifested major problems in terms of compatibility of data and interpretation as attempts to provide cross national comparison even within Europe have highlighted (Andreff, 1997; Andreff, with Bourg, Halba, & Nys, 1995; Jones, 1989). De Bosscher and colleagues acknowledge this in their own study ‘...comparisons of sport expenditure are challenging as expenditure definitions and sport policy delivery mechanisms vary considerably from nation to nation’ (De Bosscher et al., 2015, p. 109). The data relating to contribution of financial support (Pillar 1) are therefore at best an approximate measure of financial input calculated on the basis of heterogeneous definitions and operationalisations of financial investment in sport. Similar issues of functional comparability arise when attempting comparison of data relating to other pillars, such as pillar 3, participation.

Comparative research also requires consideration of sample equivalence (Øyen 1990; Ebbinghaus, 2005; Jowell, 1998; Schuster, 2007). The critical factor to consider here is whether national systems are alike or differ in significant ways. Most comparative sport policy studies to date have chosen to adopt a ‘Most Similar Systems Design’ (MSSD) (Ragin, 2006), which involves comparing key features that are different amongst similar countries while controlling for a dependent variable. This is partly due to scholars being interested in comparing their own country with others, but also because they recognize that fairly similar

countries are more likely to be comparable since they might operate in similar types of context (Lijphart, 1971).

The selection of cases for inclusion in SPLISS 2.0 is based on a number of criteria, but broadly speaking inclusion criteria represent pragmatic considerations. For example, although there are a wider number of cases involved in SPLISS 2.0, the inclusion of new countries is based on the availability and willingness of researchers or teams within these countries who have access to relevant data relating to the elite sport system in their own country. In SPLISS 2.0, “any nation interested was invited to participate under the condition that they were able to collect the comprehensive data set and follow research protocol” (De Bosscher et al., 2015, p. 66). Thus, the nations incorporated in the SPLISS 2.0 study do not include countries which have won no medals at all in Summer Olympics (35% of all nations competing) or Winter Olympics (78%) nor does it include the strongest countries in terms of overall performance. Australia is included and has the best medal performance per head and per GDP of all countries for the Summer Games, but GB is excluded, because, although it participated in SPLISS 1.0, it feared that its market position in terms of understanding Key Success Factors, might be eroded if competitor nations gained access to its own intelligence on such matters (though Northern Ireland is included).

The decision to select a sample of nations on a pragmatic basis, whilst convenient, runs the risk of selection bias or ‘problem of contingency’ (Ebbinghaus, 2005). In the case of SPLISS and other large-N comparative studies ‘selection bias’ should be acknowledged and its implications accounted for.

The pragmatic nature of sample selection means that the SPLISS framework is unable to account for some of the unique system features of the ‘powerhouse’ medal-winning nations, such as, for example, the role of the education system in the US, or the centralized co-ordination role of the Chinese government in producing elite athletes. In addition,

although there has been some shift with regard to less economically developed nations from SPLISS 1.0 to SPLISS 2.0, the SPLISS nation sample continues to focus predominantly on developed nations and hence the SPLISS account has far less utility in relation to less economically developed nations.

A final set of problems, which are linked to equivalence, relate to the inter-observer and study reliability of the SPLISS account. The reliance on local researchers in SPLISS 1.0 and 2.0 within each national context brings with it potential difficulties. This is a strategy for increasing the reliability of local data collection since as these researchers are able to assist the consortium in ensuring, for example, construct equivalence of concepts employed – some of which are evidently not appropriate and/or applicable in certain national contexts. The SPLISS consortium seeks to control for such difficulties by providing a protocol for data definition and collection. Nevertheless, the practical realities of data collection is still carried out by an individual researcher or a small team who are responsible for assessing their own nation's sporting system, including the social and political environment, sporting history and culture. The researcher(s) is / are also required to translate their own observations and data into inventories of observations in a centrally devised framework provided by the consortium. This process of standardization of data leaves considerable leeway both for researcher bias, and for the suppression of cultural specificity in individual sport policy systems.

Individual researchers and research teams participating in the SPLISS projects are self-funded. This is a practical limitation in itself in that involvement in the project is only possible for individuals with the necessary economic and social capital to be able to invest the time, and to access appropriate data required. This practical resource limitation may, in part, explain why the SPLISS nation samples are still dominated by western-northern nations despite the increase in sample size between SPLISS 1.0 and 2.0. It is also important to acknowledge the motivations of funding bodies and their limitations in relation to access to



data. For example, De Bosscher and colleagues note that perhaps the major motivation of funders for participating in the SPLISS studies is to gain greater knowledge of critical success factors in order to obtain a market advantage in sporting performance. However, the public availability of data and findings from the SPLISS studies means that this market intelligence is open to all, and where winning medals is a zero-sum game the national sporting interests of participating nations may be damaged by the ability of their rivals, whether or not they have funded a study as part of the SPLISS project, to access this information. As De Bosscher and her colleagues note, this factor explains the withdrawal of UK Sport's support for British involvement in SPLISS 2.0, having been a principal actor in the SPLISS 1.0 consortium (De Bosscher et al., 2015, p. 67).

De Bosscher and colleagues have attempted to mitigate the above-mentioned issues of equivalence. Such mitigation strategies include utilizing European-wide secondary data sets (e.g., Eurobarometer survey), data triangulation, generating detailed work protocols, defining key terms/concepts, and translating their survey instruments into multiple languages. Nonetheless, the employment of these strategies, while welcome in *mitigating* the effects of such (non-)equivalence issues, does not mean that such problems have no negative impact on inter-nation comparison (Jowell, 1998).

### **Problem 5: Overlooking/Ignoring Agency**

The selection of the unit of analysis is integral to any social research endeavour, but is particularly important for comparative researchers. Selection of the unit of analysis, the major entity being studied, is fundamentally rooted in the researchers' ontology and epistemology, as selection largely depends upon what sorts of knowledge are sought, and thus what methodological approaches can be used to gather or construct such knowledge. It is here that De Bosscher and colleagues take a particularly clear-cut stance in that they focus upon the meso-level arguing that macro-level factors "are relatively stable and cannot be influenced

significantly by direct human influence such as policymakers” (De Bosscher et al., 2015, p. 37). We would argue, contrary to De Bosscher and colleagues, that it is not analytically possible to separate the meso from the macro or micro level. Rather, we adopt a view more closely aligned to what Ragin (2014) refers to as a *comparativist* approach in that we believe it is important to engage with and define macro-entities which influence the policy process. In contrast, the approach taken by De Bosscher and colleagues can be described as a non-comparativist, in that they deliberately choose to ignore such broader notions as abstractions that need not to be operationalized. We therefore agree with the remarks of Jowell (1998) who states that comparative researchers “...should pay as much attention to the choice and compilation of aggregate-level contextual variables as they do to individual-level dependent and independent variables” (p. 197) and that “all too often, analysts seem to compare national data sets in vacuo” (ibid).

The inherent danger of the non-comparativist approach is an over-emphasis on structuralist explanations of elite sport policy, which simultaneously overlooks/ignores the influence and impact of broader social, political, and cultural (macro-level) factors on the one hand and does not account for the role of individual actors (micro-level) on the other. This is perhaps unsurprising given that the adoption of a variable-orientated approach results in the identification of variables (critical success factors) which by their nature are structural. It is these structural variables that are then correlated to derive empirical generalizations about policy domains and it is these generalizations which become explanations of elite sport success. The explanation put forward is therefore one of resources i.e. the ‘have’ and ‘have not’ nations rather than necessarily one of competency i.e. how effectively such resources are utilized.

### **Problem 6: Reductionism and Mixed Methods**

The SPLISS studies claim a mixed-methods approach in that they utilize both qualitative and

quantitative data collected through an inventory and stakeholder surveys. In the SPLISS 2.0 study, the former (inventory) involved 96 CSFs divided into 212 open-ended and closed questions across the nine pillars. The latter (survey), involved primary quantitative measures on a 1-5 Likert scale. These instruments are used to generate composite indicators and critical success factors through a mixture of qualitative and quantitative data. We take issue with the use of composite indicators (CI), which is a quantitative or qualitative measure derived from a series of observed 'facts' (Freudenberg, 2003). More recently, De Bosscher (2018) has acknowledged a number of problems with adopting a composite indicator methodology to compare elite sport policies. However, while there may be a rationale for developing quantitative composite indicators based on qualitative data in quantitative research, such a strategy should not be confused with developing a mixed methods approach. Mixed methods approaches do not develop qualitative methods and subsequently translate such data into quantitative measures. A mixed methods approach implies the subjecting of qualitative data to *qualitative forms of analysis* (e.g. subjecting documentary or interview data to qualitative forms of analysis such as thematic analysis, analysis of discourse, or ethnographic content analysis).

The subjecting of qualitative data to quantitative analysis will almost invariably involve a significant loss of meaning and thus of explanatory power, and if it does not, the value of collecting such qualitative data at all is likely to be questionable. SPLISS 2.0, for example, incorporates a large number of interviews with key stakeholders, e.g. athletes, performance directors, coaches etc. Much of the data produced was recorded in a manner amenable to quantitative analysis (e.g. the question "do athletes receive nutritional advice?" generated a response which was reported as a simple dyadic/binary (yes/no) response) and that the potential for contributing to understanding was much diminished. This type of approach can hardly be said to be the product of mixed method inquiry.

Rather than invalidating the SPLISS account, we see qualitative analysis as a particularly promising area that has the potential to significantly enhance the SPLISS account of elite sport. Not only do both qualitative and quantitative data provide complementary insights into the policy process in general but the amount of effort that De Bosscher and colleagues have expended on collection of qualitative data might be expected, given qualitative, rather than quantitative, analysis of the data, to generate some real insights into policy throughputs in the development of elite sport policies both within and across nations.

## **Conclusions**

The critique of the SPLISS approach to analysing the success of elite sports policies which we have sought to develop herein, does not seek to deny the contribution to policy explanation and evaluation which variable-based approaches to analysis can make. However, what we have identified are the limits of such analysis in generating explanations of policy success. Our claim is that SPLISS, as an analytic or conceptual framework, is useful but that its application as a model (for example in terms of prediction of performance) is at best limited and at worst inappropriate.

We have noted difficulties associated both how the SPLISS approach has been applied and the nature of the SPLISS approach *per se*. In relation to the former, the adoption of pragmatic or convenience sampling is, for example, fraught with difficulties since this requires a comparison of sporting nations of very different histories, cultures, resources and perspectives on sport, such that comparison is in principle undermined. In particular in analysis of elite sport policy success, if some measure of medal winning is the basis of constructing the dependent variable as a marker of success, then the efforts made to effect success by very many nations which have had no medal success at all, and /or whose sporting targets are reduced to a level which is more realistic for them, would render this type of dependent variable an insensitive measure of elite sport policy success. In reality, the world's

sporting nations do not constitute a sufficiently homogeneous population to foster meaningful analysis for all nations, and if meaningful generalisations are to be made, a more helpful approach would be to undertake some form of cluster analysis, grouping together like, or akin, nations (in sporting and in broader terms) and undertaking separate analyses for each cluster.

In relation to the appropriateness of the SPLISS approach in general, the potential of the approach to produce explanation of the causal links between independent and dependent variables is limited. The attribution of causes of policy outcomes cannot rely solely on statistical analysis of quantitative data relating in particular to inputs, throughputs and outputs. Detailed qualitative accounts of policy development and implementation will potentially provide explanations of particular actors', or groups of stakeholders' perspectives and explanations of how, why, and under what circumstances, particular policy actions have (or have not) proved effective in achieving desired policy outcomes. Evaluating the validity (or more accurately the warrantability of such claims, will allow us to identify and assess explanations of causes of policy change in groups of nations. In addition, it will allow us to identify and evaluate explanations within a given case or cases (i.e. within a national system or systems) of policy success. Such explanations may be unique to particular contexts, thus addressing at the individual case level (or even within the case at the individual sport level) the question of what works, for whom, under what circumstances etc. in line with the aspirations of realist policy analysis (Pawson, 2006). This argument is consistent with other comparative methodologists (e.g., Øyen, 1990, Jowell, 1998), who have argued that "comparative research may have to shift its emphasis from seeking uniformity among variety to studying the preservation of enclaves of uniqueness among growing homogeneity and uniformity" (Sztompka, 1988, p. 215). Analysis at the level of individual cases can also help

to address issues of equifinality in which different policy configurations achieve similar outcomes.

While we recognise the attempts made, particularly in the shift from SPLISS 1.0 to 2.0, to develop a mixed methods approach, we would suggest that the application of the term mixed methods to the approach adopted is mistaken. What in essence is undertaken in the SPLISS usage of the ‘qualitative’ data, is a quantitative representation of the qualitative data which is then subject to quantitative analysis. For a truly mixed methods approach to have been developed, qualitative data, rather than being converted to a quantitative format, would have been subjected to qualitative analysis. This, however, is only possible where treatment of qualitative data, whether it is in within-case or across-case analysis, is more clearly tied to formal protocols for extraction of themes, development of support for arguments, or identification of the nature of discourses etc. Such data will be particularly helpful in unpacking explanations of throughput in which the nature of the relationship between context, causal mechanisms and outputs is dynamic, given the open systems nature of policy relating to social contexts. The use of formal protocols (thematic analysis, discourse analysis, ethnographic analysis etc) will assist not simply in the generation of types of explanation, but also in relation to claims about how consistent, coherent or exhaustive, such explanations are in relation to the matters under evaluation.

The hegemony of the SPLISS approach in the field of analysis of the factors critical to elite sport policy success has been very evident in the mobilisation of an active ‘industry’ of SPLISS papers and associated conferences and workshops. Alternative, and in particular qualitative, forms of comparative analysis in the sport field have subsequently been neglected. Our paper has sought to highlight the nature of the insights that the SPLISS approach has already generated, or has the potential to generate, but also by implication to

highlight what the SPLISS approach is not able to tell us, and therefore what other forms of analysis might promise, with a view to addressing the imbalance in the current literature.

## References

- Andersen, S., & Roglan, L. (2012). *Elite Sport: Same Ambitions, Different Tracks*. Copenhagen: Copenhagen Business School.
- Andreff, W. (1997). Sports Financing in Europe - Towards a Transformation in the 21st Century. In (Vol. News Articles Section, Play the Game Website). Copenhagen: Play the Game.
- Andreff, W., with Bourg, J. F., Halba, B., & Nys, J. F. (1995). *Les enjeux économiques du sport en Europe: financement et impact économique*. Paris: Editions Dalloz.
- Beach, D., & Brun Pedersen, R. (2013). *Process-Tracing Methods: Foundations and Guidelines* (Kindle Edition ed.). Ann Arbor, Michigan: University of Michigan Press.
- Bergsgard, N. A., Houlihan, B. M. J., Mangset, P., Nødland, S. I., & Rommetveidt, H. (2007). *Sport Policy: a comparative analysis of stability and change*. Oxford: Butterworth-Heinemann.
- Busetti, S., & Dente, B. (2017). Using process tracing to evaluate unique events: The case of EXPO Milano 2015. *Evaluation*, 23(3), 256-273.
- Chalip, L., Johnson, A., & Stachura, L. (Eds.). (1996). *National Sports Policies: an International Handbook*. Westport, Conn.: Greenwood Press.
- Chen, S., & Henry, I. (2019). Assessing Olympic legacy claims: Evaluating explanations of causal mechanisms and policy outcomes *Evaluation*.  
doi:<https://doi.org/10.1177/1356389019836675>
- Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, 24(2), 119-128. doi:10.1016/s0149-7189(01)00003-9
- De Bosscher, V. (2007). *Sports Policy Factors Leading to International Sporting Success*, Brussels: VUB Press.



- De Bosscher, V. (2018). A mixed methods approach to compare elite sport policies of nations. A critical reflection on the use of composite indicators in the SPLISS study. *Sport and Society*, 21(2), 331-355.
- De Bosscher, V., Bingham, J., Shibli, S., Van Bottenburg, M., & De Knop, P. (2008). *The Global Sporting Arms Race: an International Study on Sports Policy Factors Leading to International Sporting Success*. Oxford: Meyer & Meyer.
- De Bosscher, V., De Knop, P., & van Bottenburg, M. (2009). An analysis of homogeneity and heterogeneity of elite sports systems in six nations. *International Journal of Sports Marketing & Sponsorship*, 10(2), 111-131.
- De Bosscher, V., De Knop, P., Van Bottenburg, M., & Shibli, S. (2006). A conceptual Framework for Analysing Sports Policy Factors Leading to International Sporting Success. *European Sport Management Quarterly*, 6(2), 185-215.
- De Bosscher, V., De Knop, P., van Bottenburg, M., Shibli, S., & Bingham, J. (2009). Explaining international sporting success: An international comparison of elite sport systems and policies in six countries. *Sport Management Review*, 12, 113–136.
- De Bosscher, V., Heyndels, B., De Knop, P., van Bottenburg, M., & Shibli, S. (2008). The paradox of measuring success of nations in elite sport. *Belgeo: Revue Belge de Geographie*, 2, 217-234.
- De Bosscher, V., Shibli, S., van Bottenburg, M., De Knop, P., & Truyens, J. (2010). Developing a Method for Comparing the Elite Sport Systems and Policies of Nations: A Mixed Research Methods Approach. *Journal Of Sport Management*, 24(5), 567-600.
- De Bosscher, V., Shibli, S., Westerbeek, H., & Van Bottenberg, M. (2015). *Successful Elite Sport Policies: an International Comparison of the Sports Policy Factors Leading to*

*International Sporting Success (SPLISS 2.0) in 15 Nations*. Maidenhead: Meyer and Meyer

Della Porta, D. (2008). Comparative analysis: Case-oriented versus variable-oriented research. In D. Della Porta & M. Keating (Eds.), *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective* (pp. 198-222). Cambridge: Cambridge University Press.

Digel, H., Barra, M., Bosel, J., Fahrner, M., Kohl, M., & Utz, A. (2002). *Organisation of High Performance Athletics in Selected Countries: Final report for the International Athletics Federation*. Tübingen: University of Tübingen

Digel, H., Burk, V., & Fahrner, M. (2006). *High Performance Sport: an International Comparison*. Tübingen: Brauer.

DiMaggio, P., & Powell, W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, (1983)

Dowling, M., Legg, D., & Brown, P. (2018). Cross-Comparative Sport Policy Analysis and Paralympic Sport. In Brittain, I., & Beacom, A. (Eds.), *The Routledge Handbook of the Paralympic Games* (pp. 249–272). London: Routledge

Dowling, M., Brown, P., Legg, D., & Grix, J. (2018). Deconstructing comparative sport policy analysis: assumptions, challenges, and new directions. *International Journal of Sport Policy and Politics*. 10:4, 687-704

Ebbinghaus, B., 2005. When less is more: selection problems in large-N and small-N cross-national comparisons. *International Sociology*, 20(2), 133–152.

Freudenberg, M. (2003). Composite Indicators of Country Performance: A Critical Assessment. *OECD, Science Technology and Industry Working Papers* 16: 1–24.

Green, M. (2004). Power, policy and political priorities: Elite sport development in Canada and the United Kingdom. *Sociology of Sport Journal*, 21(4), 376-396.

- Green, M., & Houlihan, B. (2004). Advocacy Coalitions and Elite sport policy change in Canada and the United Kingdom. *International Review for the Sociology of Sport*, 39(4), 387-403.
- Green, M., & Houlihan, B. (2005). *Elite sport development: Policy learning and political priorities*. London: Routledge.
- Green, M., & Oakley, B. (2001). Elite sport development systems and playing to win: uniformity and diversity in international approaches. *Leisure Studies* (20), 247-267.
- Henry, I., Amara, M., Al-Tauqi, M., & Lee, P. C. (2005). A typology of approaches to comparative analysis of sports policy. *Journal of Sport Management*, 19(4), 480-496.
- Houlihan, B. (2009). Mechanisms of international influence on elite sport policy. *International Journal of Sport Policy*, 1(51-70).
- Houlihan, B. (2013). Theorising the analysis of sport policy. In I. Henry & L.-M. Ko (Eds.), *Handbook of Sport Policy* (pp. 11-22) London: Routledge.
- Houlihan, B., & Green, M. (2008). *Comparative elite sport development : systems, structures and public policy*. Amsterdam: Elsevier ; Oxford : Butterworth-Heinemann.
- Jones, H. G. e. (1989). *The Economic Impact and Importance of Sport in Europe: a European Study*. Strasbourg: Council of Europe
- Jordan, G. B. (2010). A theory-based logic model for innovation policy and evaluation. *Research Evaluation*, 19(4), 263-273. doi:10.3152/095820210x12827366906445
- Jowell, R. (1998). How comparative is comparative research?. *American Behavioral Scientist*, 42(2), 168-177.
- Kay, A. (2005). A critique of the use of path dependency in policy studies. *Public Administration & Society*, 83, 553-571.
- Landman, T. & Carvalho, E. (2017). *Issues and methods in comparative politics* (4<sup>th</sup> edition). London: Routledge.

- Lijphart, A. (1971). Comparative politics and the comparative method. *American Political Science Review*, 65(3), 682-693.
- Oakley, B., & Green, M. (2001). The production of Olympic champions: International perspectives on elite sports development systems. *European Journal for Sport Management*, 8(1), 83-105.
- Øyen, E. (1990). *Comparative methodology: theory and practice in international social research*. London: Sage
- Pawson, R. (2006). *Evidence-based Policy: a Realist Perspective*. London: Sage.
- Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. Leeds: University of Leeds.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: SAGE.
- Pawson, R., & Tilley, N. (2004). Realist Evaluation. *Evaluation*, 48, 1-36.
- Pawson, R. D., & Boaz, A. (2004). *Evidence-based policy, theory-based synthesis, practice-based reviews*. Swindon: Economic and Social Research Council.
- Peters, B. G., Pierre, J., & King, D. S. (2005). The politics of path dependency: Political conflict in historical institutionalism. *Journal of Politics*, 67, 1275-1300.
- Petticrew, M & Roberts, H. (2005). *Systematic Reviews in the Social Sciences: A Practical Guide*, Oxford: Wiley-Blackwell.
- Ragin, C., (2006). How to lure analytic social science out of the doldrums: Some lessons from comparative research. *International Sociology*, 21 (5), 633–646.
- Ragin, C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies*. Oakland: University of California Press.
- Riordan, J. (1999). The Impact of Communism on Sport. . In J. Riordan & A. Kruger (Eds.), *The International Politics of Sport in the 20th Century* (pp. 48-66). London: Taylor and Francis.

- Riordan, J. (Ed.) (1978). *Sport under Communism*. London: Hurst.
- Riordan, J., & Jones, R. (1999). *Sport and Physical Education in China*. London: Taylor and Francis.
- Sabatier, P. A., & Jenkins-Smith, H. C. (1993). (Eds.), *Policy change and learning : an advocacy coalition approach*, Boulder, Colo. ; Oxford: Westview Press.
- Schuster, J. M. (2007). Participation studies and cross-national comparison: proliferation, prudence, and possibility. *Cultural trends*, 16(2), 99-196.
- Shibli, S., & Bingham, J., with Henry, I. (2007). Measuring the Sporting Success of Nations. In I. Henry & Institute of Sport and Leisure Policy (Eds.), *Transnational and Comparative Research in Sport: Globalisation, Governance and Sport Policy* (pp. 61-81). London: Routledge.
- Sztompka, P. (1988). Conceptual frameworks in comparative inquiry: divergent or convergent?. *International Sociology*, 3(3), 207-218.
- Wood, L., & Kroger, R. (2000). *Doing Discourse Analysis: methods for studying action in talk and text*. London: Sage.
- Yamamoto, M. (2009). *The influence of non-domestic factors on elite sport development and anti-doping policy: The cases of Japan and UK/England (Unpublished PhD thesis)*. Loughborough University, Loughborough.

## Reference

- ARCHER, M., BHASKAR, R., COLLIER, A., LAWSON, T. & NORRIE, A. (eds.) 1998. *Critical Realism: essential readings*, Abingdon, Oxon: Routledge.

- PAWSON, R. 2006. *Evidence-based policy : a realist perspective*, London, SAGE.
- PAWSON, R. 2009. Middle Range Realism. *European Journal of Sociology / Archives Européennes de Sociologie*, 41, 283-325.