Measuring Policy Learning: A framework for Measurements Development

Bishoy Louis Zaki,
Department of Public Management and Governance, Faculty of Economics and Business Administration.
University of Gent, Belgium.
Corresponding Author: Bishoy.Zaki@ugent.be

Ellen Wayenberg,
Department of Public Management and Governance, Faculty of Economics and Business Administration.
University of Gent, Belgium.

Bert George,
Department of Public Management and Governance, Faculty of Economics and Business Administration.
University of Gent, Belgium.

Presented to:
Panel T11P05: Measurement challenges in public policy research, International Conference of Public Policy (ICPP 5), Barcelona, July 2021

Note: This paper presents a preliminary draft of ongoing work. Please consider as such.

Please do not cite, circulate, or use without an explicit written consent from authors.
Measuring Policy Learning: A framework for Measurements Development

Abstract
Developing and utilising measurements is of instrumental value for policy research and practice. However, the research agenda on measurement in public policy in general, and policy learning in specific has been somewhat stymied. In this article, we focus on the development of measurements for policy learning as one of the burgeoning concepts in contemporary public policy research and practice. We argue that the policy learning measurement agenda is challenged by entwined conceptual, characteristic, and methodological issues endemic to policy learning research and practice. Thus, these issues strike at the heart of the measurement endeavour as an epistemological, methodological, and practical endeavour. To address this concoction of issues, we propose a 3-phase theoretically coherent, empirically grounded framework for development policy learning measurements that spans Theoretical Conception, Instrument Development, and Instrument Deployment. We nuance each phase with key considerations from extant policy learning and policy process literature. We conclude with a brief discussion and a set of recommendations for adapting our proposed framework to a wide range of contexts.

Key Words: Policy Learning, Measurement, Measurement Development, Public Policy, Framework
1. Introduction

The notion of measurement is central to both public policy research and practice. The availability of robust and reliable measurements of concepts can assist in knowledge acquisition, replication, and theory development (Rammstedt & Bluemke, 2019; Prewitt, 2010; Crepaz & Chari, 2017). It can also provide insights to the design quality of policies (e.g., Fernández-i-Marín et al., 2021). Measurements are also quintessential for public policy as practice. Questions such as “how much?”, “How far?” and “to what extent?” and similar others are ever so pressing in contemporary public policy. Answering these questions can give clear indications to policymakers and the public, not only as to policy performance, but also to policy options, political and ideological dispositions as well as the current state of affairs. There are two main reasons for why this can be critical. Politically, the public’s expectations from governments are increasing. These expectations and the perceptions of government performance eventually influence citizen satisfaction and political voice (Oliver, 2011). Instrumentally, contemporary policymaking is progressively technically complex, ambiguous, and multidimensional. Policy systems are also becoming growingly complex, with multiple stakeholders, competing interests, and sensemaking frames (Head & Alford, 2013; Crozier, 2010). The intensity of such issues was recently emphasized with the emergence of global wicked existential threats such as pandemics, climate change and environmental crises (e.g., Zaki & Wayenberg, 2020; Rietig, 2018; Baekkeskov, 2016). This complexity and increased public scrutiny have rendered the need for evidence-based or evidence informed policy making undeniable, despite raging debates around its constraints and utility (e.g., Sanderson, 2002; Head, 2010). Whether in a bid to fine tune, adjust or overhaul policies (Head, 2010), or to legitimate policy action (e.g., Weiss, 1986; Dunlop, 2017); evidence-based policymaking thrives on clear, measurable, comparable, and reliable indicators (Tamtik & Sá, 2012). Practical implications and reactions to public policy embodied in robustly measurable public policy concepts and indicators loopback to theory, and in doing so grounding, maturing, and extending it through drawing on clear and robust empirical observations (Prewitt, 2010).

However, though critical for theory and practice; clear, and robust measurements in the realm of public policy have been substantially lacking, an issue that has ailed the field for decades. In this article, we heed the call for empirically grounded theoretical enhancements on
measurements of concepts in public policy research. We particularly focus on the field of policy learning. Our motivation to focus on policy learning is threefold. First, academic, and practical interest in policy learning is burgeoning (Bakır, 2017). Policy learning is being continuously employed to strengthen and extend the larger study of policy theory (e.g., Dunlop & Radaelli, 2013; Dunlop & Radaelli, 2017). Second, policy learning provides substantial value for enriching policy analysis (for an overview of potential avenues, see Dunlop & Radaelli, 2020; Dunlop & Radaelli, 2018). This theoretical lens is practically employed across a range of critical fields of practice from agriculture and biotechnology (e.g., Dunlop, 2017), to pandemics and public health (e.g., Baekkeskov, 2016; Zaki & Wayenberg, 2020), to nuclear energy and environmental policies (e.g., Nilsson, 2006). The use of policy learning in empirical cases can be critical to enhancing the much sought-after evidence-based policymaking (Wagner & Ylä-Anttila, 2018), particularly as a practice employed to address commonly complex and challenging policy issues (e.g., Dunlop & Radaelli, 2016; Lee & Van de Meene, 2012). Third, despite the wide of use of policy learning, and its palatability for theory and practice, its measurement agenda leaves a lot to be desired (Radaelli, 2009), with scarcity of studies and a rather limited agreement among scholars on how it can be measured (Nowlin, 2020; Moyson, 2016).

In this article, through critical reflection on extant literature, we argue that the development of policy learning measurements is underpinned by entwined conceptual, methodological, and characteristic challenges. Policy learning literature shows key issues with conceptual clarity, cohesion, the relationships between conceptualization, operationalization, and research designs (e.g., see Dunlop, 2020; Maggetti & Gilardi, 2016; Radaelli, 2009). To address these issues, we propose a framework for policy learning measurement development, that is grounded in policy learning theory, practice, and social sciences measurements literature. This framework spans the multifaceted nature of measurements, that is: epistemological, methodological, and practical (Bruschi, 2017). To achieve this endeavour, we are guided by three central questions: What is the current state of measurement in policy learning research? What are the central challenges in developing policy learning measurements? How can these challenges be systematically addressed?
In answering these questions, we provide policy learning scholars with a theoretically grounded 3-phase framework for developing policy learning measurements that considers the inherent characteristics of policy learning theory and empirical research. We also provide practitioners with a toolkit for scrutinizing and evaluating existing policy learning solutions and designing new, nuanced policy learning measurements. This article proceeds as follows. In section 2, we provide an overview of the current state of measurement in policy learning research and shed light on the three categories of challenges, conceptual, characteristic, and methodological. In section 3, we propose a framework for developing policy learning measurements and in section 4 we conclude with a discussion and conclusions.

2. The current state of measurement: Conceptual, Characteristic and Methodological challenges

In this section, we shed light on the current state of measurements in policy learning research and the key challenges faced on such an endeavour. However, before delving into the current state of affairs, it is of relevance to reflect on the emergence of policy learning as a formal field of study. The notion of learning in public policy emerged decades ago (See: Deutsch, 1966), mainly as a supplementary perspective to enable a better understanding of policy action and policy change. At that time, these issues were mostly understood through the lenses of powering or vested interests (Hall, 1993; Biegelbauer, 2016). Thus, policy learning emerged as an ontologically and epistemologically complex concept that spans various theoretical and empirical standpoints, attempting to explain entwined forces and phenomena. From there on, interest in policy learning has significantly grown leading to a “branching out” of the field into various subfields. This field’s growth sprout has been quite divergent, in a manner by which the analytical utility of the field and its theoretical status has been subject to interesting debate (e.g., Dunlop & Radaelli, 2018; Goyal & Howlett, 2018). With this understanding in mind, what is it that could be said about the state of measurement in policy learning?

Despite the centrality of policy learning in contemporary policymaking and existing attempts to measure facets of policy learning, the measurement agenda has been largely stymied and measurements remain rather scarce with relatively limited consensus on operationalisation
(Radaelli, 2009; Pattison, 2018). However, it is fair to say that several scholars have embarked on different measurement endeavours whether focused on measuring policy learning or its variant concepts. This included measuring learning through the time taken for policies to be adopted (Berglund & van Warden, 2006), measuring learning intensity, consensus formation, and policy transfer (Montpetit, 2009), the degree of learning (Nedergaard, 2009), identifying teachers or exemplars (Lee & Van de Meene, 2012), identifying different types of learning versus no learning (Radaelli, 2009), the amount of belief change (Moyson, 2017), or policy diffusion (Maggetti & Gilardi, 2016). However, throughout those measurement attempts, scholars have voiced concerns over various challenges. A review of policy learning extant literature shows that such concerns fall under three broad yet highly visible categories: conceptual foundations (e.g., Nedergaard, 2009; Maggetti & Gilardi, 2016), the nature of policy learning and its contexts (e.g., Moyson, 2017; Moyson, 2016), which naturally leads to concerns over methods and research designs (e.g., Radaelli, 2009). Three issues are closely entwined in the context of policy learning. We first elaborate on those challenges before proceeding to propose a coherent solution strategy.

**Conceptual Challenges**

A field’s conceptual state significantly influences the ability and direction for developing measurements, after all we cannot reliably measure what we cannot adequately define (Adcock & Collier, 2001; Sartori, 1970). The creation of reliable measurements is predicated upon semblances of theoretical consistency, definitional clarity, and adequate representations of constructs of interest (Bandalos, 2018; Kaplan, 1964). However, the study of public policy (as the general study of social sciences) suffers tangible conceptual issues (e.g., Gerring, 1999; Carlsson, 2017; Hupe & Hill, 2006). In terms of conceptualization, though interest in policy learning has grown, such growth was somewhat fragmented and lacked consistency. The concept has grown ambiguously, with a “dizzying array of definitions” (Karlsen & Larrea, 2016; Stark, 2019), to the extent of being signposted as a “conceptual minefield” (Levy, 1994). This has long contributed to issues of conceptual cohesion, and consistency (For a relatively early account on this issue, see Bennett & Howlett, 1992). Policy learning is also commonly viewed as “conceptual umbrella” for several mechanisms, concepts, and effects (e.g., see: Anglund, 1999). This umbrella includes a
myriad of concepts such as lesson drawing (Rose, 1991), policy diffusion (Maggetti & Gilardi, 2016), policy transfer (Dolowitz & Marsh, 2002) among tens of other concepts. In many cases, the label of policy learning and labels of those concepts are used interchangeably, and in heterogeneity (Maggetti & Gilardi, 2016). The super/sub ordinance of these concepts in relationship to policy learning also varies across scholarly work. These conceptual issues are known to pose challenges for measurement creation and comparability of findings, particularly given the lack of axiomatic standardization, conceptual and theoretical consistency which leads to “linguistic- conceptual anarchy” (Bruschi, 2017), and what Sartori (1975) termed a “tower of Babel” phenomenon. Hence, these conceptual challenges obviously bound measurement and operationalization, particularly as conceptual refinement is a central presupposition to measurement (see: Adcock & Collier, 2001; Goertz, 2006; Sartori, 1970). As such, we posit that the issue of conceptual clarity can partially explain issues with policy learning measurements development, particularly: the scarcity of measurements (Moyson, 2016), the lack of agreement on operationalizations (Pattison, 2018), and notoriety of measuring facets of policy learning (Nedergaard, 2009).

These issues led to calls for more explicit reflections on the relationship between conceptualization and operationalization (e.g., Maggetti & Gilardi, 2016; Radaelli, 2009), and identifying clear criteria while attempting to measure policy learning (Dunlop, 2020). A systematic literature review we recently conducted including 147 empirical policy learning articles highlights and enumerates the conceptual ambiguity and fragmentation underlying the field (For more details on the review process, check Appendix I). Our results show that 81 articles (51%) provided no definition for policy learning, while the remaining 66 articles (49%) offered a staggering set of 34 distinct definitions. Furthermore, in terms of policy learning types, we identified a staggering set of 61 different labels indicating types of policy learning (usually each used in very low frequencies). Though these types of learning could be conceptually similar, they are presented under different labels and seem to be syntactically different (and seldom elaborately defined). Hence, it becomes highly subjective to discern whether they can be amalgamated, combined or taxonomized. Thus, as we established the necessity for conceptual
and theoretical clarity and consistency for measurement development, their current state in policy learning research shows significant potential for refinement.

**Characteristic and Methodological Challenges**

Moving from conceptual and theoretical issues, the very nature and characteristics of policy learning and its context pose significant methodological challenges to creating measurements.

Policy learning is intrinsically compound and complex (Bennett & Howlett, 1992). The practical and multidimensional nature of policy learning understandably complicates the creation of measurements (Pattison, 2018). Policy learning is a process, thus not merely a single variable to be measured (Montpetit & Lachapelle, 2017). It is also multi-actor, multi-dimensional and occurs across various systems, structures, and modes, this includes individuals, organizations, systems, and groups, often simultaneously (For examples, see: Heikkila & Gerlak, 2013; Dunlop & Radaelli 2017; Dunlop & Radaelli, 2013; Dudley, 2007). Furthermore, as a practice, policy learning is politically embedded, motivated, and moderated; with various stakeholders bringing competing interests, perspectives and sensemaking frames (e.g., Sabatier, 1993; Weiss, 1986; Zaki & Wayenberg, 2020). Policy learning is also socially and politically desired, even normatively appealing as concept in the policy process (Nilsson, 2006). Afterall, the notion of learning is naturally palatable for us as individuals, as well as critical to demonstrate for policymakers and organizations. This is given its ability to legitimize and defend policy action through making decisions (presumably) based on best weighted evidence (Radaelli, 2009; Weiss, 1986; Wagner & Ylä-Anttila, 2018), an issue that has manifested in recent crises through the “following the science” mantra used by many governments (e.g., Zaki & Wayenberg, 2020).

Methodologically, this makes identifying genuine instances of learning relatively more challenging. Many logics underly this difficulty. Policy learning is socially desirable (Moyson, 2017), thus, willingly, or unwillingly policymakers can overestimate or ignore the influence of learning when being questioned over it (Radaelli, 2009; Nedergaard, 2009). This is all the while we are generally unable to fully grasp or scrutinize the true inspirations or motives of policymakers (Legrand, 2012). Furthermore, policymakers (as some of the main objects of measurement, whether directly through interviewing or indirectly through crafting policy
documents) could also be unaware they actively engaged in a process of policy learning (Nedergaard, 2009).

An additional methodological challenge that cannot be underestimated is the temporal one. Policy learning does not happen (or at least manifest) overnight, thus there is a general preference for analysing learning over long periods, usually a decade or so (e.g., Nilsson, 2006; Fendt, 2010). Thus, while looking at extended periods it is almost impossible not to find instances of learning as Radaelli (2009) rightfully proclaims. This means that we run the risk of being overinclusive in our analyses or while identifying the policy learning phenomenon. Last but not least, while there are literature inclinations to methodologically identify or measure policy learning through observing policy change. However, in many cases policy learning takes place without observable or immediate changes, and policy change can occur driven by other factors such as power shifts, without genuine policy learning taking place (Nilsson, 2005).

Piecing together the conceptual, theoretical, characteristic, and methodological challenges endemic to policy learning research shows they are fundamentally entwined and interconnected. It also shows that this concoction of issues spans the entire spectrum of the measurement development endeavour, that is epistemological, methodological, and practical as shown in table (1).

<table>
<thead>
<tr>
<th>Measurement Facets</th>
<th>Relevance to Policy Learning Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measurements as epistemology</strong></td>
<td>Limited Conceptual Cohesion, Conceptual Fragmentation and ambiguity (e.g., Borrás, 2011), Theoretical Consistency (e.g., Goyal &amp; Howlett, 2018), Multidimensionality (e.g., Pattison, 2018).</td>
</tr>
<tr>
<td><strong>Measurements as Methodology</strong></td>
<td>Objects of learning unwilling or unaware of learning or overestimate learning (e.g., Radaelli, 2009; Nedergaard, 2009).</td>
</tr>
<tr>
<td><strong>Measurement as Practicality</strong></td>
<td>Temporal dimension complicates identification of learning (e.g., Radaelli, 2009; Nilsson, 2005).</td>
</tr>
</tbody>
</table>

Table (1): Policy Learning challenges across measurement facets
As we established the reach of policy learning challenges across key facets of measurements, in the next section, we proceed to offer a systematic framework aiming to alleviate some of those challenges.

3. A framework for policy learning measurement development

While challenges undermining the development of measurements in policy learning are tangible, the utility and value of measurements are lucrative, thus making concerted efforts to facilitate measurement creation a worthy feat. Drawing on policy learning theoretical and empirical works, we propose a guiding framework for measurement development. Here, we view a framework as a structure underlying a certain issue with a specific purpose (Stamer et al., 2016). More specifically and borrowing from information sciences, we argue for the utility of a framework that acts as a reusable set of abstract solutions for an observable problem. This is predicated upon three main tenets: being a sum of components and patterns, being reusable (in part or in whole), and being customisable (Johnson & Foote, 1988). We posit that such features should be central in designing problem-oriented frameworks for policy learning, particularly given the need to adapt and customize policy learning (both as a research and practice) to varying contexts (For examples on the importance of coping with variability within a policy learning context, see: Karlsen & Larrea, 2016; Fiorino, 2002) and its multidimensionality, complexity and multiplicity of components.

The proposed framework is based on a nuanced understanding of the theoretical and empirical features of policy learning research; thus, our main interest is in providing key caveats to strengthen the process of measurement development across the various facets of measurement, epistemological, methodological, and practical. However, here we must ascertain that this framework does not intend to fully incorporate or replace the entirety of measurement development processes at a foundational level. After all, this endeavour remains beyond the scope of a single article.

A broad view of measurement

First and foremost, given our relentless quest for conceptual clarity, it is of utility to first elaborate on what we mean by measurement. For the purposes of this section, we view measurement as “a process linking abstract concepts to empirical indicants” (Zeller & Carmines,
Here, we sidestep (while acknowledging) the narrower view of measurements in social sciences as a process of mathematisation or assignment of numbers to objects, originally rooted in the natural sciences perspective (e.g., see Stevens, 1951; Kaplan, 1964). A broad view of measurement can be particularly helpful given the nature of policy learning research, that is predominantly qualitative and focuses on phenomena that are not necessarily quantifiable through conventional enumeration.

**A Phased approach to measurement creation**

In section 2, we identified several challenges for measurement development in policy learning, particularly *conceptual, characteristic, and methodological*. We argued that those challenges are inherently entwined and interdependent. Though scarce, some articles measuring policy learning (or some of its facets) have attempted to address some of those issues. For example, Moyson’s (2017) focus on addressing cognitive biases, social desirability and Radaelli’s (2009) focus on alleviating measurement errors, enhancing conceptual analysis and assessing the null hyposthesis. However, a concerted and systematic effort to address the aforementioned entwined issues across epistemology, methodology and practical deployment is yet to be seen.

To address those challenges, we synthesize and adapt a 3-phase framework for policy learning measurement development spanning *Theoretical conception, instrument design* and *deployment*.

Our proposed framework traverses the multifaceted nature of measurements in social sciences: epistemological, methodological, and practical (Bruschi, 2017). It also links and spans the three key challenge areas we identified in section two (concepts, characteristics, and methods) through a theoretically coherent approach. This is by emphasizing the relationships between measurements and: conceptual structures (Goertz, 2006; Maggetti & Gilardi, 2016; Gerring, 1999; Adcock & Collier, 2001), system’s, characteristics, and contexts systems (e.g., Zeigler et al., 2019; Hick, et al., 2019) and methods (e.g., Mathison, 1988; Patton, 1999; Denzin, 1978). We nuance each of the proposed phases with key considerations by drawing on policy learning extant literature and theory. In doing so, we attempting to systematically address some of the main challenges that underpin validity and reliability of policy learning measurements. In designing this framework, we avoid being normative or overly perspective, thus allowing for it to
be used for a wide range of approaches to measurement. Hence, we intend that this framework is used as a customizable guiding process rather than a specific set of steps valid for an overly limited range of cases. Next, we move to elaborate key considerations for each of the framework’s phases.

**Phase I: Theoretical Conception**

The theoretical conception phase is mainly concerned with scrutinising and consolidating the theoretical and conceptual foundations that underpin measurement development. Clarifying and succinctly articulating conceptual foundations presuppose measurement development and enhance validity (Adcock & Collier, 2001). This process goes beyond conceptual statements, into engaging in a process of critical reflection on the relationship between conceptualisations, underlying structural criteria and measurement (e.g., Maggetti & Gilardi, 2016; Dunlop, 2020; Radaelli, 2009; Zeller & Carmines, 1980). In a policy learning context, this requires careful choices to be made, particularly given the findings of our systematic review indicating that varying (and sometimes conflicting) conceptualisations for policy learning exist. Here, what we term as the “theoretical conception phase” spans three recursive steps concerning the identification of: Basic, secondary level and indicator conceptualisation as proposed by Adcock & Collier (2001), and extended by Goertz (2006) and Maggetti & Gilardi (2016). We elaborate on each of them in sequential order.

First, theoretical conception starts with a focus on clearly delineating the central “basic level concept” underlying measurement development. Goertz (2006) presents the basic level concept as most important from a theoretical perspective. Given the nature of policy learning, we argue that such process of delineation should result in a conceptual statement that meets some criteria of conceptual goodness, particularly ones that are relevant to- and enables valid measurement (e.g., such as those leading to the identification of core structural dimensions of a concept). An example of such criteria was offered by Gerring (1999) being: coherence, differentiation, depth, and theoretical utility. Furthermore, the existence of different conceptualisations with the same label in policy learning can be troublesome (see: Borrás, 2011), yet it is understandable given that policy learning draws on a wide range of interdisciplinary resources and ontologies (see: Dunlop & Radaelli, 2017). In such cases, resulting concepts can
have different theoretical anchors and ontological purposes (see: Ansell, 2019). These anchors, ontological presuppositions and orientations should be carefully considered when identifying a basic level conceptual structure while observing the intended use of the target measure. Here, as an example, we argue for the utility of our earlier proposed conceptualisation of policy learning as “the circulation and consumption of policy issue related information and knowledge among actors in a policy system and structure, within a policy context.” Mainly given its ability to act as a “background concept” that leads to a systematised conceptualisation through reflecting on the research goals. In this context, a background concept (as the one we put forward) includes the potentially diverse meanings associated with policy learning as the phenomenon in question (Adcock & Collier, 2001). This does not mean that the aforementioned conceptualisation should be a normative standard, however, it is offered as an example of a structured background concept that can be functional across a range of cases. Building on the centrality of the background concept suggests that a clear articulation of the theoretical lenses underlying the conceptualisation and measurement creation should be consistent.

The second dimension focuses on identifying the properties of the concept to be measured (Bruschi, 2017). This should involve a critical reflection on conceptual entailments, in other words what a concept structurally is (Goertz, 2016; Ansell, 2019). This largely corresponds to Goertz’s conception of the “Secondary level”, which Maggetti & Gilardi (2016) articulate as pertinent to the “constitutive dimensions” of the concept. In the context of policy learning, this entails reflecting on what properties of learning are we attempting to measure? From this level forward, one might consider disposing of “measuring policy learning” being an overinclusive statement, particularly as policy learning is a complex process, rather than a single variable to be measured (Montpetit & Lachapelle, 2017). Here, various approaches to identifying such properties at the secondary level exist. For example, Maggetti and Gillardi’s application of the Goertz’s hierarchy of levels in policy diffusion was from a somewhat mechanistic view, thus offered learning, emulation, and competition at the secondary level. For the purposes of this article, we adopt a different perspective, perhaps viewed as mostly “characteristic” given that our conceptual focus is on policy learning. Here, literature shows that there are several properties of learning that can be measured. For example, the extent of learning (Moyson, 2017), intensity
of learning (Montpetit, 2009), the longevity of learning outcomes (Corbett et al., 2018), the inclusiveness of learning sources (Zaki & Wayenberg, 2020), openness or potential for learning (Nedergaard, 2009) or the influence of certain tools on learning (Radaelli, 2009).

This takes us to the final step of the theoretical conception phase, that is identifying indicators able to faithfully represent the secondary level concepts or properties identified. Drawing on policy learning literature, we argue that indicators should consider two main criteria. First, the ability to identify the full range of the secondary level states, thus in a manner can be used for falsification or identification of the null condition as a key aspect of measurement (For example, see Radaelli, 2009 and Nedergaard, 2009). Second, validity of indicators in terms of capturing the conceptual dimensions sought through measurement while avoiding omissions or excessive distortion or an exacerbated risk of capturing confounding phenomena (Adcock & Collier, 2001). This step lends further focus to our earlier emphasis on the importance of conceptual goodness and critical interrogation of conceptual entailments as primary steps in the theoretical conceptual phase. Thus, the need for conceptual frameworks with clear and disentangled dimensions. This is particularly important as the clarity of concepts lends to the sound identification of the properties to be measured, which a key presupposition of the next phase, that is instrument development (Bruschi, 2017; Sartori, 1970).

By the end of the theoretical concept phase, four main questions should be answerable within a certain degree of confidence, those include: Do we have a theoretically coherent and sound conceptualisation of what to be measured? Is the theoretical lens employed consistent with our conceptualisation and measurement context? Do our indicators capture the properties we aim to measure without distortion or omission? How sensitive are our indicators to the full range of property states? An overview of the theoretical conception process is illustrated in figure (1).
Phase II: Instrument Development

The instrument development phase is concerned with the creation of “agents” capable of assigning empirical indicants to the indicators being measured. Typically, as agents; measurement instruments facilitate the collection of data on which empirical or hypothesis tests can be conducted (Leplège, 2003). Inherently, the instrument of measurement comes to contact with an object of measurement, thus is able to measure or gauge the state of property in question manifested by said object. As a rudimentary example, if the range of interdisciplinary experts involved in a policy learning process is a property of policy learning inclusiveness (See: Zaki & Wayenberg, 2020), an instrument should enable gauging the state of said property. However, an emphasis needs to be made here. The instrument’s focus should naturally be on the state of manifestation or property, rather than the object of measurement, which can naturally possess a complex and almost infinite set of properties (Bruschi, 2017). In many cases, the objects of measurement (e.g., actors) are involved in extremely complex processes, systems, and
interdependent relationships (See: Crozier, 2010). While this issue is not inherently straightforward under best conditions, it can be even more challenging in policy learning research. Policy learning has been shown to happen within a complex, and even messy policy process with layers upon layers of cross interactions (e.g., Wilson, 2019 Dunlop & Radaelli, 2017), where objects of measurement are deeply entangled with - and discursively influence the properties we attempt to measure. This entanglement is not only political or organizational, but micro foundationally cognitive and intellectual (For examples see: Kamkhaji & Radaelli, 2017; Heikkila & Gerlak, 2013). To grasp this complexity, we need to take a step back and reflect on a central aspect of policy learning, that is its multidimensionality and multifaceted nature. To do so, reflecting on decades of policy learning research, we draw a paradigm from the study of quantum physics, namely the phenomenon of quantum entanglement. Paradigms are based on scientifically concluded findings and are able to address “residual ambiguities” and induce valid predictions through valid inferences (Kuhn, 1996).

Simply put, the notion of quantum entanglement applies to studying the of most fundamental particles of existence. At the foundational level, particles within our universe exist and behave in dual forms (a particle and a wave at the same time with uncertainty in simultaneously identifying where and what they are). Taking this view to policy learning is by no means a far stretch, next to quantum entanglement being a universal fact of existence; reflection on policy learning literature shows that it indeed expresses itself as in such entangled state. In our proposed conceptualisation of policy learning, we showcased that the salience of multilevel interactions between actors, systems, structures, information, and contexts in a continuous state of flux. In this state of flux, there are varying degrees of each of those components. For example, at any point in time actors can exist and be viewed as individuals (e.g., Moyson, 2017), individuals within collectives (e.g., Heikkila & Gerlak, 2013), organisations or institutions and systems (e.g., Stark & Head, 2019), at the same time advocates or brokers (Howlett et al., 2017). Also, in almost all cases, actors can have multilevel interactions (Dunlop & Radaelli, 2017). Hence, also in almost all cases, it is likely that indicators identified would entail measurements of several properties in superimposed states, i.e., investigating entangled properties that reflect a consolidation of multiple forces, roles, influences, and states (e.g., discourses, rules, beliefs, biases, etc.). This
issue was generally articulated for social science measurement by (Bruschi, 2017) as objects manifesting potentially an infinite number of states. For example, with a key informant (be it a policymaker), in an ideal (and hardly achievable) case, measurement instruments should be able to disentangle (to a certain extent) different properties and states of said agent. This includes a policymaker’s belief entrenchments (e.g., Sabatier, 1988), cognitive biases and the social desirability underlying learning (e.g., Moyson, 2017), influence of institutional mandates and rules (Marier, 2009), or even shrouds of secrecy surrounding certain policy issues or policymaker inspirations (e.g., Legrand, 2012). The same issue extends to different instruments such as archival analysis of policy documents, meeting minutes, etc., all with the potential of compromising both the instrument’s validity and reliability.

While we cannot argue for a single global solution for these issues, we argue that two main approaches at the instrument design level can alleviate some of their burdens. Namely, system’s modelling, and triangulation. First, we draw on the notion of system’s modelling from system’s theory. Systems’ modelling aims at describing or mirroring complex systems across multiple views, mainly in an attempt to understand how those systems function under different conditions (Hick et al., 2019). Such an approach is potentially valuable given the growing complexity, interconnectedness, and nonlinearity of policy systems in which learning is embedded (e.g., Crozier, 2010; Crozier, 2008). Under the conceptual umbrella of policy learning, formal deductive modelling can result in improved specificity in the identification of variables, which is much needed (Evans, 2009). According to Zeigler et al. (2019) System’s modelling theory draws two main distinctions, mainly between the inner system constitution (structure) and manifestation (behaviour). This perspective of systems’ theory acknowledges two central concepts: decomposition (as breaking a system down into component systems), and composition (as coupling component systems to form larger systems). This approach as Zeigler et al., (2019) frame it, leads to somewhat of a relational hierarchal view. One central adaptation its application within a policy learning context entails emphasizing that systems modelling goes beyond a sum of parts. Here, once again consolidating conceptual foundations (as elaborated under the theoretical conception phase) offers significant value as a presupposition of instrument development. For example, if we are to adopt our proposed conceptualisation of policy learning
as an exercise of ontology, core features of said conceptualisation (background and systematic) can offer insights to the modelling approach. In this case, drawing on the conceptualisation emphasizes that the interactions between components are multilevel, complex, and multidirectional. Thus, modelling the learning environment’s behaviour needs to consider the relationships and logics of interaction among its various components. Policy learning literature has shown that nuancing structures with logics of interaction and relationships can yield varying system behaviours and outcomes (e.g., Dunlop, 2017a; Dunlop & James, 2007; Dunlop, 2017; Zaki & Wayenberg, 2020). Thus, adopting a system’s modelling approach can contribute to more in-depth understanding of potential structural influences or forces acting on the object of measurement, limitations, and key considerations for the objects with which instruments come into contact. This can better inform the instrument design process on how to manoeuvre or potentially isolate such influences towards better validity and reliability. It can even be through re-aligning the identification of objects of measurement through carefully plotting potential biases or issues.

Additionally, employing a systems modelling approach prior to instrument development can also lend a helping hand to one of the most emblematic challenges of policy learning; measuring or assessing the relationship between learning and change. A relationship that is complex and “analytically blurred” (Borrás, 2011; Fenger & Quaglia, 2016). Throughout literature, there is an evident inclination to consider policy change as an indicator for learning (Raudla et al., 2018; Fenger & Quaglia, 2016), despite the fact that change can happen without learning (Birkland, 2006). Thus, by using a system’s modelling approach, researchers can model system specific change mechanisms and configurations. In doing so, garnering insights into the design of appropriate instruments to detect/test for learning along those mechanisms with an informed focus on key dimensions (An example in our case includes: actors, structures, information and knowledge, and contexts). This can lead to a more reliable assessment to whether learning has influenced key actuators of change within a system and as such, to what extent has learning contributed to change. Semblances of systematic modelling in policy learning (yet at more specific levels such as actors) have been shown to yield better understandings of major changes or reforms (e.g., Bandelow, 2017; Dunlop & James, 2007).
Our second proposition pertains to triangulation, understood as the use of multiple methods or data sources to develop more comprehensive understandings of phenomena (Patton, 1999). First, we propose that employing mixed methods to formulate more comprehensive measurement instruments can allow for disentangling the multiple superimposed layers underlying policy learning measurements, particularly given the value mixed methods offer for understanding complex phenomena and scrutinizing obtained results which can enhance validity (Molina-Azorin, 2016; McKim, 2015; Miles & Huberman, 1984). Here, our proposition is not by any means normative, rather can be employed at the researcher’s discretion whenever needed and with multiple permutations (e.g., using two different quantitative or qualitative methods or a combination of qualitative and quantitative methods).

So far, the use of mixed methods in policy learning in general has not been so frequent. The results of our systematic review show that only two articles out of 147 clearly demonstrated the use of mixed methods, which certainly leaves a lot to be desired. Second, employing instruments that observe principles of data source triangulation can also assist in strengthening measurement instruments. Naturally, triangulation of data sources should contribute to the validity of findings (particularly based on measurement of phenomena), by showing a convergence of evidence (or at least lack of contradiction). This is particularly important in qualitatively dominated inquiries, where subjective biases are naturally more common (Mathison, 1988), an issue acknowledged in policy learning research where the very basic identification of learning involves substantial subjective judgements (Birkland, 2006). However, the value of triangulation lies not exclusively in its confirmatory function, triangulation can assist in scrutinizing measurement results towards either a better understanding of their meaning or their invalidation. The notion of triangulation that we put forward here does not necessarily fall shy of data sources and methods but can also extend to other forms of triangulation such as those of investigators, theory, or persons (Denzin, 1978). Furthermore, Denzin’s ahead of its time understanding of triangulation also involves triangulation across time space, an aspect critical to bring forward in the case of policy learning measurements development where temporal and spatial dimensions influence the process of learning, its dynamics, and outcomes (e.g., Mooney & Lee, 1999; Nilsson, 2006; Dunlop, 2009).
Last but not least, a proliferating issue of measurements in social sciences is that scholars tend to develop measurements that are highly case specific, with either ambiguous or very specific uses. Beyond their use in almost exclusive cases, such instruments are rendered of almost no utility, an issue that also draws on splintered conceptual positions (Bruschi, 2017). This case is quite evident in policy learning literature, where to our knowledge no measurement developed has been used more than in a single instance. Hence, we propose that substantial focus should be placed upon a clear and robust articulation of the measurement instrument assumptions. An articulation of instrument assumptions delineates the conditions and contexts where the measurement instrument can be used, an issue that is problematic in social sciences measurements in general, and policy learning measurements in specific (Adcock & Collier, 2001). Here, the instrument assumptions should be somewhat consistent with the underlying theoretical and ontological assumptions of the concepts at the center of the measurement endeavor (For an excellent example see: Moyson, et al., 2016). We argue that articulation of instrument assumptions should be structured around the key dimensions underlying instrument design, which once again circles back to the importance of conceptual goodness, and structural clarity.

Figure (2): Overview of the Instrument Design Phase
The conclusion of the instrument development phase involves a discursive reflection on the theoretical conception phase where the consistency of ontological and conceptual assumptions are rechecked to ensure that the transition from abstract conceptualisation to instrument development has not compromised the instrument’s underlying conceptual consistency. Here, rephrasing two main questions posed by Adcock & Collier (2001) on the validity of measurement the instrument can act as a litmus test, namely: have key elements been omitted from the indicator? Does the measurement include superimposed “distortions” that can compromise its validity?

**Phase III: Deployment**

In this phase, we define deployment as using the instrument within a certain environment for measurement. Our focus here is two dimensions, temporal and spatial, Thus our main concern is the choice of “when” and “where” the measurement instrument is used.

Policy learning scholarship has long acknowledged the temporal perspective. For example, the seminal Advocacy Coalitions framework looks at policy-oriented learning on a preferential time scale of a decade or more (Sabatier, 1988; Feindt, 2010; Sabatier, 1993). This time-sensitivity (particularly with attention to long periods) to empirical observations of learning has been somewhat consistent in literature (e.g., Dunlop, 2009; Moyson, 2016). After all, actors can take years to infuse their learning into policy mechanisms and policy action (Fiorino, 2002). While this relatively long timespan can provide enriching insights, it can also pose burdens (Dunlop, 2009). Chief of which is that looking at such long periods will almost always lead to finding instances of learning (Radaelli, 2009). This further warrants the our earlier focus on clearly identifying central concepts and types of learning being measured, thus excluding potential distortion in the measurement process and avoiding the indentification of misleading instances of learning. So far, this is only an issue of ‘time’ in measuring learning. Reverting to our earlier examples of properties of learning to be measured (e.g., Longevity, inclusiveness, intensity, etc.) requires the consideration of the time needed for the identified properties to manifest or vary.

While time has been a major concern for policy learning measurement, literature largely overlooks the issue of “timing”. In other words, are we measuring learning at the right timing? And is the measurement instrument timing-appropriate? When measuring learning, we usually
look at different sources of information, ranging from key informant interviews to documentary analysis (e.g., Weissert & Scheller, 2008; Radaelli, 2009). Such sources are naturally highly time sensitive. For example, literature shows that in many cases, actors involved might not want to acknowledge they have learned something (Nedergaard, 2009), an issue very plausibly timing and context contingent. This is particularly a conceivable notion, given that learning is politically infused (Biegelbauer, 2016). Also, while learning can take time to brew, its manifestations (represented in policy action) are unlikely uniform across time (e.g., Fiorino, 2002; Nilsson, 2005). For example, according multiple streams framework, learning might not be articulated in the policy arena unless certain conditions are met or policy brokers emerge or the policy issue survives agenda-setting, despite learning brewing below surface (e.g., Hudson & Kim, 2014).

Furthermore, even when learning is seemingly identifiable, it is argued that in some cases early reactions to policy issues (particularly crises) might not necessarily constitute genuine instances of learning (e.g., see Kamkhaji & Radaelli, 2017). Again, this emphasizes the importance of clear links between background and systematic concepts, indicators, and modelling systems behaviour and structures with focus on measurement agent positions as earlier discussed. As in the first and second phases, the deployment phase also concludes with a discursive relfection guided by three central questions here: Has the appropriate amount of time elapsed for the properties in question to materialise? Is the timing for instrument deployment opportune in a manner that leads to reliably ascertaining the state of property being measured? Last, is the instrument in place functional given the contexts said timing? With that said, to ensure, theoretical, conceptual and methodological consistency, these reflections should loopback to the theoretical conception and instrument design phases. In doing so, adjusting either re-adjusting or replacing theoretical assumptions or instrument development in light of the instrument deployment context. An overview of the Instrument Deployment phase is depicted in figure 3.
4. Preliminary Discussion and Conclusion

Measurements are a central and sought-after endeavours in contemporary public policy. Despite their appeal, the agenda on measurements has been somewhat lacking. A key challenge for this deficiency can be found in the conceptual and theoretical state of the field, a landscape viewed by Hupe & Hill (2006) as mountain islands of theoretical structure with limited cohesion and shared methodological underpinnings. Policy learning is by no means an exception, as a matter of fact it is one of the most emblematic embodiments of these challenges in the study of public policy, particularly when its complex and multidimensional nature is added to the mix. In this article, we mapped out the theoretical, characteristic, and methodological challenges endemic to policy learning across the multiple facets of measurement; epistemological, methodological, and practical. We then proceeded to offer a 3-phase theoretically coherent, empirically grounded, process-oriented systematic framework that attempts to address those
challenges. Put together, this framework consistently links theoretical, conceptual, and ontological foundations and assumptions underlying measurement with instrument design and deployment as shown in figure 4.
Figure (4): An Integrated Framework
In designing this framework, we observed two key criteria underlying the design of frameworks, being flexibility and reusability. We view that beholding these criteria is central if we are to address issues in policy learning. Particularly with the fields varying ontological “dividing lines” (e.g., Dunlop & Radaelli, 2017), levels of complexity (e.g., Bennett & Howlett, 1992; Pattison, 2018) and multidimensionality (e.g., Pattison, 2018; Dunlop & Radaelli, 2017). This issue of contexts is one of utmost salience in policy learning research. Contextual configurations and factors can significantly influence policy learning processes and outcomes (e.g., Karlsen & Larrea, 2016; Dunlop & Radaelli, 2013; Zaki & Wayenberg, 2020). Naturally, this implies that such factors influence the facets of policy learning we aim to measure, and the agents that measurement instruments interface with. This process dynamizes the identification of key considerations and caveats for policy learning measurements development.

Here, we emphasize that while our proposed framework attempts to address key issues in designing policy learning measurement, it does not aim to offer a full stock of issues or present an exhaustive listing of considerations that works across all contexts. However, it proposes a general outline and a logical structure for integrating considerations across varying contexts and research designs. Thus, while we argue that the higher level phases of theoretical conception, instrument development, and deployment are valid within different contexts; considerations within those phases should be open for variation. However, we also argue that the governing logic for adding or removing key considerations is predicated upon safeguarding two central logics observable in measurement development literature. First, the consistency of epistemological, ontological, and conceptual aspects on one hand, and their link with methodological design and deployment contexts on the other. This consistency allows for the development of coherent measurement instruments that are better positioned to capture policy learning phenomena with validity and reliability. Second, the articulation of design assumptions. Such articulation can avail critical knowledge about the ontological, methodological and contextual determinants of measurement instrument validity. Thus, it allows for the effective use of such instruments within appropriate contexts. It also allows for an informed customisation of instruments to be repurposed under various conditions.
Bibliography


