

# Evaluating policy measures to tackle undeclared work: the role of stakeholder collaboration in building trust and improving policy-making

---

GREY Working Paper No. 8

Sheffield University Management School  
University of Sheffield

Anton Kojouharov and Colin C Williams

### **Acknowledgements**

This Working Paper is an output of the European Commission's Framework 7 Industry-Academia Partnerships Programme (IAPP) grant no. 611259 entitled 'Out of the shadows: developing capacities and capabilities for tackling undeclared work in Bulgaria, Croatia and FYR Macedonia' (GREY). The authors would like to thank the funders for providing the financial support to enable the research and writing of this paper. The usual disclaimers apply.

## Contents

1. Executive Summary.....	4
2. Introduction.....	5
3. Policy evaluation - overview.....	6
3.1. Why evaluate.....	9
3.2. Evaluation approaches and models of research uptake.....	11
4. Challenges and dilemmas.....	17
4.1. Evaluating measures to tackle the grey economy – some lessons.....	27
5. Toward a genuine collaborative approach – bridging the gap of institutional asymmetry.....	33
6. Conclusion.....	38
Bibliography.....	40

## 1. Executive Summary

The aim of this paper is to examine and analyse the realm of policy evaluation approaches and methods as they relate to assessing measures to tackle undeclared work. The discussion is set at the backdrop of a brief review of the more prominent theoretical and conceptual considerations in the policy evaluation literature. The paper then investigates results from policy assessments and evaluations illuminated in the previous GREY working papers, as well as some selected from the Eurofound database. The analysis of a limited sample of available policy evaluations and results demonstrates that a common probable cause of policy failure with regards to measures to tackle undeclared work in the GREY countries is an apparent lack of knowledge and understanding of the needs, attitudes and abilities of policy beneficiaries and targets. This strongly suggests non-existent or inadequate engagement of all stakeholders in the initial policy evaluation or feasibility study stage of the policy-making process. This exploratory effort has identified the adoption of a collaborative stakeholder strategy to tackling the undeclared economy as a major recommendation for moving forward.

## 2. Introduction

This working paper is part of a series of deliverables within the project entitled ‘GREY – Out of the shadows: developing capacities and capabilities for tackling undeclared work<sup>1</sup> in Bulgaria, Croatia and FYR Macedonia’ (‘GREY’), funded by the European Commission’s Framework 7 Industry-Academia Partnerships Programme (IAPP). The objective of this project is to provide concrete policy recommendations about what policies may work better in tackling the undeclared economy in two new EU Member States (Bulgaria and Croatia) and a candidate country (FYR Macedonia)<sup>2</sup>. In particular, the objectives of this project are to:

- Conduct a comprehensive review of the policy approaches and measures being used to tackle undeclared work.
- Empirically evaluate the measures for facilitating the shift of undeclared work into the declared economy in the selected target region.
- Based on the results of the first two objectives, to recommend possible new policy approaches and measures (or endorse old ones) that will facilitate the shift of undeclared work into the declared economy.

The aim of the present Working Paper No. 8 is to provide an overview on some of the challenges in the methods and approaches to policy evaluation in the field of tackling undeclared work. In doing so, this paper will build upon the research findings in the baseline assessments (Working papers No. 1-3) as well as Working Paper No.6, which provide valuable insights into evaluating policy transferability (see Williams et al., 2014; Dzhekova and Williams, 2014; Dzhekova et al., 2014; Franic and Williams, 2014). As such this attempt will not be making iconoclastic claims and propositions, but rather, in true working paper spirit, it seeks to set in motion a meaningful and informed discussion about a complicated and controversial issue, namely policy evaluation as it relates to measures to tackle the undeclared economy. In doing so, it poses questions and exposes issues which can only be addressed by further research.

This paper will refrain from offering a comprehensive overview and discussion of the various known methods and approaches to evaluating programs<sup>3</sup>, as this has been explored in depth elsewhere (Nonchev, 2009; Mathison, 2005). Moreover, the reviewed literature suggests that there is no one right evaluation solution, as approaches and methods depend largely on context. Instead, the aim of this working paper is to review the developments in evaluation theory and practice and to consider their consequences for tackling the undeclared economy in the target countries. Furthermore, extrapolations from evaluation theory and practice will be deliberately selective, given that there is a particular social and political context present in our focus states, namely - high levels of undeclared work, a prominent asymmetry between formal and informal institutions, and an authoritarian political and administrative legacy. In the field of tackling undeclared work, the broader context is further identified to include an over-reliance on top-

---

<sup>1</sup> The terms undeclared work, grey economy, informal economy will be used interchangeably

<sup>2</sup> Throughout this paper Bulgaria, Croatia and FYROM will be collectively referred to as the GREY countries.

<sup>3</sup> For the purpose of this paper the terms policy, program, measure and intervention will be used interchangeably.

down positivist/quantitative research for policy-making and evaluation, which often fails to capture the actual policy needs, attitudes and abilities of stakeholders and intended policy targets and beneficiaries. This paper puts forward a tentative argument that collaborative stakeholder involvement in policy evaluation and feasibility consultation minimises the adverse effects of the risks of policy failure inherent in the contextual policy environment of the GREY countries. Moreover, stakeholder-centric collaborative strategies in themselves carry a value-forming potential capable of altering context in favour of institutional trust-building, perceptions of administrative fairness and accountability, therefore moving toward bridging the gap of institutional asymmetry.

The paper is structured as follows. Following this Introduction, section 2 provides a conceptual overview of developments in the field of policy evaluation research up to the present day. Section 3 exposes some of the challenges and controversies surrounding policy assessment and reviews by situating it within context. Section 4 presents some practical experiences from evaluating measures to tackle the undeclared economy, drawing upon findings from the baseline assessments and from practitioners and experts collected through the Eurofound database. In doing so, we aim to synthesize and analyse the more pertinent difficulties evaluators and policy-makers encounter when reviewing and drawing lessons from implemented measures. Section 5 then suggests and examines a collaborative stakeholder evaluation approach in line with current trends in policy-making and governance.

### 3. Policy evaluation - overview

The ascent of policy evaluation into the mainstream arena of governance is associated with developments in the US in the 1950's and 1960's, when social research came to be viewed as a viable evidence producing tool that informs policy-makers (European Commission, 2013). A strong positivist movement was posed to link the scientific method with power, politics and policy-making (Aelbek, 1995). A similar approach took height in the New Labour governance philosophy of the late 1990's in the UK as an integral ingredient of the push for evidence-based policy-making. The UK Modernising Government document of 1999 states:

“This Government expects more of policy makers. More new ideas, more willingness to question inherited ways of doing things, better use of evidence and research in policy making and better focus on policies that will deliver long-term goals” (Cabinet Office, 1999: 16)

This shift of linking science and evidence with power and policy making aimed at arming government with a rational and realist approach to policy making and design that focused on the “what works” paradigm (Cabinet Office, 1999; Solesbury, 2001). The efforts of the New Labour government toward an evidence-based policy making approach brought about intensified institution-building that was to incorporate and implement the new direction of governance philosophy. These included: the National Institute for Health and Clinical Excellence; the Social Care Institute for Excellence; the Migration Advisory Council among others. Valuable experience, insight and guidelines to evaluating policies are summarised in two essential

documents - HM Treasury's *The Magenta Book - Guidance for Evaluation* and the *Green Book - Appraisal and Evaluation in Central Government* (HM Treasury, 2011a,b)

In the EU, the drive toward evidence-based policy-making and evaluation is largely related to the requirements of the cohesion and structural funds in the processes of harmonisation, standardisation, expansion and development (European Commission, 2013). The 2001 European Commission White Paper on Governance maintains that:

“Scientific and other experts play an increasingly significant role in preparing and monitoring decisions. From human and animal health to social legislation, the Institutions rely on specialist expertise to anticipate and identify the nature of the problems and uncertainties that the Union faces, to take decisions and to ensure that risks can be explained clearly and simply to the public.” (European Commission, 2001: 19).

EU expansion and structural fund beneficiary requirements have been regarded as largely responsible for “importing” evidence-based practices and evaluation into Southern European states, as well as into post 2004 member states (European Commission, 2013a:11). Evaluation aims and responsibilities are clearly formulated at every stage of the programming cycle (ex-ante, interim, and ex-post). Starting in 1988, the structural funds evaluation approach has evolved considerably. One of the main changes in the evaluation approach in the EU has been to transition from externally imposed evaluation responsibilities to internally stimulated demand for evaluation studies stemming from policy-makers and program officials. In addition, the emphasis of the evaluation effort moved away from over-relying on evaluating at the end of the policy cycle and towards a more integrated approach that focuses on including evaluation research at the beginning of policy design (European Commission, 2013a: 12)

For the current funding period (2014-2020), a paradigmatic shift had been envisioned by a significant revision of regulations and guidelines on program evaluation. Perhaps most crucial and pertinent is that the result of policy is no longer to be measured solely as statistical change in indicator variables, which may or may have been influenced by the specific measure. Rather, particular attention should be focused on whether and how the measure has produced change. EU wide guidelines have been published in most funding areas (see European Commission 2013b). The Commission has outsourced most evaluation research to external organisations in an attempt to increase perceptions of objectivity and accountability. Moreover, the Commission's directorates have been lacking the required resources to conduct evaluations internally. The evaluation approach and guidelines of the Commission are not only important for improving policy-making at the EU level, but they also provide expertise and guidance to local authorities where policy evaluation has not been organically embedded in the policy cycle (Hoerner and Stephenson, 2013). Despite the fact that policy evaluation at the EU level was introduced and institutionalised much earlier than in many member states, the topic remains under-researched. On the one hand, there is a proliferation of published evaluation reports. On the other hand, however, very little attention has been focused on studying the context within which evaluation is conducted and the factors that influence the evaluation process and findings (Hoerner and Stephenson, 2013).

The development of national and EU-level evaluation policies and approaches is indicative of a diverging and disparate evaluation environment across policy areas and jurisdictions. This is significant. There appears to be a rift in the level of institutional expertise and experience with evaluation across Europe, wherein established democracies such as the UK, France, Germany, the Scandinavian states, among others, have enjoyed an organic development of evaluation policies, while newer EU member states have adopted such approaches in complying with EU funding regulations and conditions. This discrepancy, among other factors, has shown the need for a framework for policy transfer analysis, as it has been established that states and institutions may seek to adopt and adapt an already tested and evaluated program from a different social and jurisdictional context (for a discussion see Williams et al., 2014). Apart from the futility of transferring measures that are strongly embedded in a social and economic context, which will be discussed later in this paper, program transfer poses risks to the effectiveness of the policy cycle as well. As Pawson (2007) maintains, it is people that make programs work – policy experts – whose tacit and professional expertise may be argued to be also, more or less, a product of a specific cultural and institutional context. Therefore, transferring, adopting and adapting ready-made measures, too, carries an inherent risk of disrupting the policy cycle by omitting the particular policy design expertise.

In this paper, we bring context and its significance for the policy cycle to the forefront. It is widely recognised that most of the “classic” literature on policy evaluation is American, followed by a surge in UK authors, especially after the change in political philosophy of the New Labour government in the late 1990’s and the arrival of evidence-based policy-making at the forefront of the UK government agenda (Albaek, 1992). In contrast, particularly in Eastern and South-eastern Europe, policy evaluation was not an organically grown element of political decision-making and the policy-cycle in general. Therefore we proceed with caution in extrapolating findings and making assumptions as we are faced with diverging contexts in which evaluation occurs. In doing so, we are inevitably drawn to the realist method of policy evaluation in asking “what works, for whom and in what context” (Pawson, 2002, 2003, 2007). We will further attempt to frame the context within which evaluation is to occur and propose practicable approaches that are appropriate in either fitting within the contextual reality and/or designed with the intent of mitigating risks to efficient evaluations that are inherent in the context itself. As policy evaluation is also a policy measure/tool, a more ambitious task is to propose approaches to policy evaluation that have the potential to improve those elements endemic in a context that contribute to the predominance and severity of the policy problem itself – in our case – a context of asymmetry between formal and informal institutions and low tax morale that perpetuates non-compliant activities. This determination is grounded in institutional theory, more precisely in the belief that policies, together with their tools and mechanisms, are value-loaded/normative means of communication between institutions/government and citizens, therefore having the power to shape social context (Lascoumes and Le Gales, 2007).

### 3.1. Why evaluate

The rationales in favour of policy evaluation are often found at the intersection of public and governmental interests. The two most readily identifiable reasons for policy evaluation research on a global scale are accountability and better and more informed policy making (European Commission, 2013; Cabinet Office, 1999; Magenta). Indeed, these are more often than not the reasons given by governments in support of evaluation research utilisation. In societies where the growth of evidence-based policy has been organic, the reasons for an increased utilisation of policy evaluation have been stated to lie in the emergence of the modern managerial state (Jasanoff, 2004), the recognition of policy subjects as policy/service consumers (Palfrey, 1992), an increased attention to policy-relevant scientific knowledge by policy-makers with a purpose of legitimizing policies of interest in a transparent and democratic society (Ezrahi, 2004) and the modernist-rationalist drive of developed OECD states to reform the public sector employing a results-oriented management paradigm (Sanderson, 2002). In contrast, in less developed and newer EU-member states for example, policy evaluation has been imposed as an external requirement for accountability, particularly as conditionality for receiving EU funding, which has been viewed as crucial for their development (European Commission, 20013; Hoerner and Stephenson, 2013).

#### *Public accountability*

In simplest terms, policy evaluation yields evidence that government is working effectively (Sanderson, 2002). It is argued, however, that public accountability in established democratic societies is to be understood in persuasive terms grounded in practicality of choice and revision rather than in normative, constitutional or moral presets (Palfrey, 1992). In this sense, policy evaluation provides for an engaged public accountability, wherein policy evaluation evidence is intended to inform program change, enhancement, extension etc., rather than serve as a simple governmental signpost. Public accountability, through accessible evaluation research, fosters a democratic constructive dialogue between policy-makers, policy targets and the wider public in a collaborative manner that instils trust and confidence in the policy-making and enforcing governmental bodies (Palfrey, 1992).

*Inform the policy-makers.* Evaluation, as a systematic performance review provides an opportunity of grasping a much broader picture of the overall effectiveness of implemented policies. For politicians, this enhanced knowledge could potentially be utilised to improve chances of election and/or re-election. Policy evaluation is often used for political purposes as policy-makers may seek to legitimise their actions veiled in the authority of scientifically produced evidence (Ezrahi, 2004; Weiss 1979; Rist 1995; Albaek, 1995, Leicester, 1999). For public officials, evaluation results contribute in understanding the cost-effectiveness of implemented programs, which would assist in enhancing strategic and operational plans for further interventions.

Sanderson (2002) argues that there are two main forms of evidence required in this approach to facilitating and increasing governmental effectiveness. The first is the evidence which promotes accountability by way of illuminating the particular results of a policy – evidence that government is working effectively. The second is evidence applied in improving public service

through more effective policies and programmes – evidence of how well they work in different circumstances. These two forms of evidence are different in substance and purpose. The first is predominantly in the form of information on features of performance and is evident in the growth of performance management in public institutions. The second form of evidence is qualitatively different from the first, as it delivers knowledge of how policy programs achieve change in society overall. Conventionally, it is assumed that reliable knowledge provides a sound basis for effective action, as it is explanatory and theoretical, providing an understanding of how policies work (Sanderson, 2002: 3).

#### *Optimising public resource allocation.*

Program evaluation helps determine whether those groups in society who experience the highest urgency for government resources and intervention actually receive them. Evaluation provides evidence about *program effectiveness* or whether goals and objectives have been met and how. Furthermore, *program efficiency and cost-effectiveness* are also illuminated through systematic evaluation (Palfrey, 1992; Sanderson, 2002; Pawson, 2003; Pawson and Tilley, 2004). Evidence on the achievement of policy objectives at the lowest possible costs and the extent to which the benefit/cost ratio is maximised are key factors, particularly in circumstances of constrained budgets, economic crisis and austerity measures. Systematic policy impact and feasibility assessments assist in mitigating the risk of ideologically loaded policy decisions or political “hunches” by evaluating whether policy objectives may indeed be achieved with available allocated resources. In this sense, policy evaluation acts as a check on unrealistic and potentially costly politically motivated policy proposals grounded in populism with the sole purpose of attracting electoral attention.

#### *Institutional learning and capacity development.*

Policy evaluation is an essential means toward enhancing public services by providing key tools and drivers for institutional learning and capacity building. This is partly due to program and performance assessment being part of a management philosophy which increasingly intends to change the behaviour of the implementing agents and then become part of managerial steering (Bjørnholt and Larsen, 2014). Organizations are often treated as cognitive systems in and of themselves, despite organizational members being capable of learning at an individual level. Hedberg (1981: 6) puts it as follows: “As individuals develop their personalities, personal habits, and beliefs over times, organizations develop their world views and ideologies. Members come and go, and leadership changes, but organizations’ memories preserve certain behaviours, mental maps, norms and values over time”. Hence, institutions learn as lessons are disseminated throughout the organization and are stored in some sort of organizational memory. In this sense, program evaluation plays a key role in the process of drawing lessons as it ideally identifies to organisational strengths and weaknesses in policy implementation and causes for policy failure. Therefore, developing and instilling an evaluation culture, as part of the overall policy-cycle, is crucial to policy improvement. Performance and policy evaluation results are intentionally intended to increase both organizational and political learning and are to be part of a learning cycle in which evidence-based knowledge guides decisions in order to improve performance in the policy cycle overall. In ideal circumstances, they feed back into policy processes, and

organizations can learn what they do well and where improvements are needed. This implies that the lessons drawn through performance measurement and policy evaluation prompt organizations and politicians to change their priorities and work processes in order to enhance policy results (Bjørnholt and Larsen, 2014). It is important to note that evaluation criteria, indicators and methods of measurement are both part of changing and/or controlling institutional behaviour. However, they also indirectly frame how political problems are being defined and delineate the perceived policy choices (Bjørnholt and Larsen, 2014).

It should be noted, however, that studies into policy evaluation research and organisational learning have increasingly demonstrated that the incorporation of knowledge into the knowledge base of an organisation is sporadic, selective and temporal (Rist, 1995: xxiii). Moreover, decision-making processes occur in an environment where rational information and evidence are simply part of a myriad of factors influencing the policy cycle. Therefore, program evaluation risks redundancy if not supported at the political level.

### 3.2. Evaluation approaches and models of research uptake

Since at its core, policy evaluation is conducted to produce information or knowledge about a specific measure, definitions of its substance, purpose and role vary widely and are, to a large extent, contingent on the epistemic convictions of the definer, i.e. positivist, constructionist or realist. A cautious formulation states that “evaluation can be viewed as a structured process that creates and synthesizes information intended to reduce the level of uncertainty for decision makers and stakeholders about a given program or policy” (McDavid and Hawthorn, 2013:3) or “the process of determining the extent to which objectives have been attained” (Steele, 1970:6).

Program Evaluation: a broader definition. A more positivist definition is found in the Magenta Book maintaining that “evaluation examines the actual implementation and impacts of a policy to assess whether the anticipated effects, costs and benefits were in fact realised, therefore also identifying what works in a program” (HM Treasury, 2011: 7). Finally, a realist evaluation investigates “what is it about a programme that works for whom in what circumstances” (Pawson et al., 2005: 22).

Indeed, diverse epistemological assumptions support the rationale behind the respective evaluation approaches and methodologies. All other features essentially derive from these fundamental perspectives. An evaluator must carefully consider the selection of the most relevant tools, methods and methodologies available that are appropriate to achieve the evaluation purpose and targets. Methodologies may utilise quantitative methods – such as macroeconomic models, input–output analyses, statistics, and control groups – and/or qualitative methods, such as surveys of beneficiaries, stakeholder interviews, and case studies (see Table 1). Approaches and methods are often determined by the direction of the evaluation research, which may rely on aggregate data gathered from secondary sources, such as national public administrations and/or regional statistical offices or data collected directly from the stakeholders involved, in order to record and identify experiences and opinions at the local level. The scope of stakeholder involvement in qualitative evaluation methods may also be varied, from engaging

only those directly participating in the programme to securing a broader, more comprehensive involvement of the community, despite being marginally affected by the policy (McDavid, J., Hawthorn L, 2013; Hoerner and Stephenson, 2013). Financial and budgetary considerations also play a role in selecting methods and methodologies, which ultimately affects the scope of the evaluation study. Last but not least, the discretion and position of the evaluator may also be a key factor in selecting a particular methodological framework - i.e. whether the expert is a neutral, impartial observer, or considered an independent variable that exerts some impact on the findings. Some take a ‘God’s Eye View’, while others consider the evaluator a policy stakeholder in their own right since he/she is ultimately a recipient of technical assistance budgets, and an agent of the principal who pays for the evaluation (Hoerner and Stephenson, 2013).

Social scientists have been engaged in profound epistemic debates on the role and nature of policy-relevant social science for decades now. In the field of evidence-based policy making and evaluation research, three schools of thought seem to have exerted the most influence in evaluation practice and literature – positivism, constructionism and realism.

*Positivism.* In line with the classical positivist traditions, it is deemed possible to obtain objective knowledge through observation (Hutcheon, 1971; Williams, 2005). Thus, positivists see policy results as the consequence of a cause and effect relationship. Methodologically, positivists are predisposed to favour quantitative methods such as econometric models, input–output analyses, and statistics. This approach strives to draw broad and comprehensive conclusions that can make generalisations and categorisation of social phenomena. This tends to support more often than not a “top-down” analysis focused on the macro-level, wherein regional or local exactitudes and contexts are deliberately overlooked. Information is gathered on an aggregate level, with individual respondents/participants considered as passive agents and there lacks a dialogue between stakeholders (McDavid and Hawthorn, 2013). This school of thought views evaluators as neutral observers, autonomous in their judgements and free from bias arising in the process of policy implementation and detached from the parties involved. A key feature, resulting from applying positivist methods, is the overreliance on numbers, as values are thought to contaminate research results and objectivity. This produces evaluation results that are void of contextualisation and fails to capture the true impact of a policy aimed at tackling a complex social problem. Context is avoided in its entirety, as exogenous factors may appear to hinder the logic of the presupposed or expected cause-effect relationships of observed phenomena. This rational theory driven system of conducting policy and evaluation research is used to support the notion of causality, for example linking program investment/cost to intervention results (Hoerner and Stephenson, 2013).

*Constructivism.* More recently, constructivist thinking has also begun exerting some influence upon evaluation approaches and methods. Constructivism in evaluation research maintains that facts and values cannot be clearly disassociated and that truth and knowledge are always constructed, even though it may not be possible to assert how (European Commission 2013, p. 21). Objective knowledge is normally precluded, as there are only subjective opinions and observations (Williams, 2005; Little, 2010; Eglash, 2011). In other words, the outcome of

evaluation is treated as resultant from the interplay of certain factors specific to a particular situation, whereby theoretical considerations are deployed to assist in understanding the external social reality (Armstrong and Wells, 2006). In this respect, constructivist evaluations employ a bottom-up approach as assumptions and social constructs of stakeholders are identified and investigated with a hermeneutic-dialectical method. Case studies are also a preferred method of generating policy knowledge. In doing so, the evaluation study places particular importance on the localised context including historical and cultural dimensions. The evaluator places particular emphasis on the local situation and its historical and cultural dimension (Armstrong and Wells 2006, p. 266). More crucially, all stakeholders participate actively in constructivist evaluation design and practice, therefore evaluators themselves are accepted as subjects of the evaluation process (European Commission 2013, p. 21). The evaluator is expected to defend his/her findings and adapt future assumptions according to what was learnt throughout the evaluation process.

*Realism.* Akin to constructivism, realism is an “umbrella” under which many different theoretical approaches across disciplines can be found and can mean different things. Pawson and Tilley (2004: 10) maintain that realist evaluation has “no preference for either qualitative or quantitative methods. The precise balance of methods to be used is selected in accordance with the realist hypothesis being tested, and with the available data”. This situates realistic approaches between positivism and relativism/constructivism, especially in terms of methods used, as no distinct preference is given between utilising quantitative versus qualitative techniques. Therefore, epistemologically, realism occupies a middle-ground between classical positivism and relativist constructivism, as realists accept different accounts and categorizations in programme evaluation (McDavid and Hawthorn, 2013). The approach is generally concerned with finding out why things work in different contexts (European Commission 2013, p. 21, Pawson, 2002a, 2002b, 2004, 2006, 2007). Realism is a theory-based approach, in that programs are regarded as theories and the evaluation is anticipated to produce theories and identify patterns in the social and organizational processes that the programme has brought about (Armstrong and Wells 2006; Pawson, 2006). The epistemic middle ground of realism, presupposes its intermediate position in the selection and use of evaluation methodologies. Evaluations make use of both quantitative and qualitative tools, whereas both top-down and bottom-up approaches depending the parameters and requirements of the particular context. Often a mix of the two approaches is preferred (Pawson and Tilley 2004, p. 10). As far as stakeholder engagement is concerned, realists envision for an active involvement, although evaluation results are steered to serve the interests of policy-makers (Bachtler and Wren 2006, p. 147). Because in realist evaluation the intervention must be re-constructed as a theory, the evaluator designs an open system, wherein he/she is in constant communication with programme participants in a teacher-learner relationship (European Commission 2013, p. 21). Therefore, policy participants’ feedback is reflected in the overall evaluation findings.

Table 1. Methodological approaches to evaluation (Davies et al., 2000:6)

Primary research						
Organised systematic empirical enquiry						
Qualitative Methods	Pluralistic Approaches	Quantitative Methods				
Focus groups, in-depth interviews, ethnographic approaches, and discourse analysis	Multi-method strategies, in-depth case studies with quantitative and qualitative data gathering  Qualitative work carried out in preparation for large-scale quantitative work  Qualitative work carried out to explore and explain quantitative findings	<i>Longitudinal Designs</i>				
		<i>Cross-Sectional Designs</i>	<i>Observational only</i>	<i>Quasi-experimental</i>	<i>Theory-driven evaluation</i>	<i>True experiments (randomisation)</i>
		Surveys of users' and providers' experience of services	Cohort studies (prospective follow-up studies), case-control studies (retrospective designs), statistical modelling techniques (econometrics)	Before-and-after studies with limited control over extent and timing of interventions	Pluralistic approaches underpinned by comprehensive theory, possibly involving quasi- and true experimentation	Random allocation to experimental or control groups
Secondary research						
Research that takes as its objects of study <i>primary</i> research studies						
<b>Systematic Review</b> – all studies relevant to a given evaluation issue are uncovered, assessed for methodological quality, and synthesised.						
<b>Meta-analysis</b> – quantitative estimates of overall effect size aggregated across a given group of primary studies.						

In addition to epistemic approaches to evaluation research, or put simply, the supply side of knowledge, various models of evaluation study utilisation have been proposed. This classification pertains to the manner, process and organisation of the actual use of evaluation results. In addition to being grounded in epistemic assumptions models of evaluation research utilisation are contingent on jurisdictional, political and social context.

*Utilization-focused evaluation.* The utilisation-focused approach proposed by Patton (1997) puts use as the driving rationale behind the model. Use “begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, from beginning to end, will affect use” (Patton, 1997: 20). The evaluation research results are therefore strictly targeted to suit intended users or stakeholders, defined as “people who have a stake – a vested interest – in evaluation findings” (Patton, 1997: 41). Overall, the evaluation process is designed with a clear intention of facilitating the use of findings by all stakeholders. It is thus crucial to identify all relevant stakeholders at the onset of evaluation study design, wherein evaluation parameters are drawn in consultation. This approach presupposed that all stakeholders have demonstrated an initial commitment to invest resources in the evaluation process, thereby requiring a certain cost-sharing dimension among end-users (Contandriopoulos and Brousselle, 2012).

*Democratic evaluation.* Unlike utilisation-focused evaluation democratic evaluation aims to provide the wider community with dependable and pertinent information to support public debates (Contandriopoulos and Brousselle, 2012). In this model, the evaluation process does not advocate for the interests of specific stakeholder groups, but serves to facilitate among opposing or competing stakeholder views and constructs (House and Howe, 1999: 96). The policy positions of all parties are duly taken into consideration and represented in the evaluation effort, wherein an independent evaluator maintains a balance of values and interests (Contandriopoulos and Brousselle, 2012). The model rests on the assumption that evaluation and policy-making takes place in a pluralistic and democratic manner, with contesting views and positions being put forward by various stakeholders. Therefore, democratic evaluation lends itself most useful in cases of highly complex, contested and/or controversial policy issues, for which there are widespread public concerns. The evaluation effort is committed to an independent and democratic assessment process and a pluralistic view is adopted by considering all stakeholder positions, thereby balancing out any power disparities the stakeholder groups may experience outside the evaluation process. In this respect the model fosters democratic debate and addresses issues of power relations and misrepresentation of stakeholders in the policy-cycle process (Contandriopoulos and Brousselle, 2012).

Weis (1979) proposes a classification of the various models by way of the intended and purposeful utilisation of public policy research. Depending on the political and social context, policy evaluation research reaches policy makers in different ways and for different reasons. In each model Weis summarizes both the rationale and the challenges that program evaluation may encounter.

*Knowledge-Driven Model* Evaluation informs policy-making directly and in a linear model, wherein scientific research and knowledge is d..... This positivist model supposes a linear process of evidence uptake directly into policy design and implementation. Based on objectivist natural science approach.

*Problem-Solving Model* - Evidence is used to fill-in a gap of knowledge preventing the design and implementation of a policy. The model works under assumed consensus between policy makers and scientist/academia on policy goals and substance of solution. In this arrangement research evidence may be either antecedent to policy issue identification or commissioned specifically to fill in an established gap of knowledge.

*Interactive Model* – Policy-makers are actively engaged in seeking information not only from social scientists, but from a diverse pool of sources, including -administrators, practitioners, politicians, planners, journalists, clients, interest groups, aides, and friends. This model is used in cases of heightened policy issue urgency, wherein policy-makers may lack the time to wait for research results and seek to justify and legitimise their judgment with reliable information from other sources. Policy evaluation research, in this case, is only one part of a complicated policy cycle.

*Political Model* – Research is utilised to support a predetermined political position driven by, for example, ideology, election cycles etc. The risk associated with this model is that evaluation becomes politicised and evidence and findings are controlled and monopolised by bureaucrats.

*Tactical Model* – In this model politicians and administrations use research to deter criticism, justify support for unpopular policies, and/or demonstrate administrative proactiveness and responsiveness. No concrete findings or evidence are presented, only the claim that research is being conducted and results anticipated.

*Enlightenment Model* – Policy-makers are influenced by “the concepts and theoretical perspectives that social science research has engendered” rather than specific findings and evidence. An additional role of social research in this model is that social research alerts decision makers to new issues and helps identify and frame new policy problems.

(summarised from Weis, 1979)

Weis's claim that the Enlightenment model of evaluation research uptake occurs most widely in practice is supported by later scholars as well. Albaek (1995) argues that the results of evaluation may have an "instrumental" or "conceptual" influence/impact/utilisation on policy and decision makers. The instrumental influence entails a more direct and immediate consideration and incorporation of evaluation findings in the decision-making process. Whereas, a conceptual influence will have a more "enlightened" effect, reaching policy-makers in bits and pieces via unsystematic channels, including media, news, journals etc. It has been, however, established that, the conceptual influence exerts a far more reaching impact on policy makers, as decision-makers may rethink and re-consider whole philosophies and approaches underlining certain policies. The instrumental effect of evaluation, on the other hand, may lead to immediate but restricted to procedures results (Albaek, 1995:8-9).

#### 4. Challenges and dilemmas

In searching for a common ground for discussion we find that a prominent converging characteristic of evaluation research and undeclared work is their deep contextualisation. On the one hand, the drivers behind non-compliant or undeclared work have recently been increasingly associated with social factors such as tax morale, public and institutional trust and perceptions of administrative and judicial fairness (Torgler and Schneider, 2009; Williams and Franic, 2015; Williams and Horodnic, 2015; Williams and Martínez, 2014; Williams and Shahid, 2015; Williams et al., 2014b, 2015b). On the other hand, most specialists agree that policy evaluation is embedded in political and social context. Moreover, policy evaluation is contingent on the overarching context of the dominant policy-cycle model in a particular society and/or jurisdiction (Weis 1979, Leicester, 1999, Pawson, 2003; Pawson and Tilley, 2007; D'agostino 1996). In this sense, for example, a particular definition of the policy issue may put policy makers in a path dependency, wherein methods and methodologies for evaluation are presupposed in the ways a problem is framed and program goals are determined.

The proliferation of social research activity since the latter half of the 20<sup>th</sup> century coupled with advances in data collection, processing and analysis tools, has made it possible for social scientists to study vast sets of regularly produced data sets, such as crime and attitudinal surveys, enabling the identification of trends in social behaviour and phenomenon over time. In this sense, context can now be measured statistically and its indicators, such as tax morale, given a value. Nevertheless, context may not be purely understood as quantified regression analysis. It needs a conceptual framework that makes sense of it. Comparative political, economic and historic studies have all propagated various explanatory models illuminating the causes and rationale in both diverging and converging political and social developments across borders and time. In many fields of the social sciences, there has been a movement away from rigid theoretical frameworks for understanding social phenomena, such as rational-actor theory and positivism and toward a more constructivist or co-productionist (see Jasanoff, 2005) view on human behaviour that takes into account the contexts out of which particular social constructs arise and within which they interact.

We therefore look for challenges and risks in the field of policy evaluation outside the technicalities of quantitative and qualitative methods and venture into the larger social, economic and political context within which it occurs. In the next section therefore, we examine some of the challenges in policy evaluation arising from context, while gearing the discussion to factors and developments pertinent to our focus societies in South Eastern Europe.

In 1999 in an influential article, Leicester succinctly outlined the risks of policy evaluation as a part of the broader concept of evidence-based policy, calling attention what he termed the “seven deadly enemies of evidence-based policy”.

**Bureaucratic logic, Cynicism and Civil service culture** – “things are right because they have always been done this way. It is a logic that reinforces traditional interpretations of policy problems by only noticing evidence that fits the existing analysis”. Leicester describes a culture of lack of trust in information gathering and dissemination, as well as the persistent support of conventional wisdom logic, despite the awareness of its falsehood or redundancy. In our case we will look to the contextualised aggravating issue of politicisation and its impact on policy cycle and organisational learning.

**The bottom line** – This challenge pertains to the over-reliance on positivist top-down approaches that produce politically exploitable numbers. Yet, as Leicester notes “these numbers, however, say nothing about the quality of services, the true test of the effectiveness of policy”. We will look at how policy problem definition and a lack of pluralistic culture of trust contribute to institutions zealousness to rely on numbers in policy evaluations.

**Consensus** – Leicester sees the danger in consensus as a justification for ignoring policy-relevant evidence. Once a consensus is reached the government sees no need to support its decisions with social science research.

**Politics** – As with Weis’s political model, political interests override evidence, risking irrational decisions. Or as D’agostino succinctly put it: “Any particular epistemology is itself a practice embedded within and meaningless outside a particular worldview. All that remains, as a basis for negotiating our common lives, is politics” (D’agostino, 1996: 398).

**Time, or the lack of it** – Administrations are often strained for resources, particularly in times of economic hardships and austerity measures. Evaluation is then often left outside top institutional priorities and is conducted only if time and resources are allocated for it.

(summarised from Leicester, 1999)

### ***Problem definition and agenda setting***

Explanations and definitions of social problems impact on the agenda-setting for policies aimed at managing them. In turn, policy agendas provide the framework for goal-setting of specific policy instruments, which ultimately contributes in defining policy performance and evaluation indicators (Dery, 1984; Rochefort and Cobb, 1995; Veselý, 2007; Hanberger, 2001). For

example, explaining the prevalence of undeclared work with a simplistic rational-actor theory model presupposes a policy choice of direct measures, such as inspections and detections, and a set of quantitative indicators for program performance and success, such as the number of control activities and value of levied fines and sanctions. On the other hand, recognising the undeclared economy as a complex and dynamic social issue calls for a balanced policy approach that includes indirect policy measures as well, thereby adding a strong qualitative dimension to measurement and evaluation techniques. Above all however, the recognition and acceptance of the undeclared economy as a “wicked” problem demands the inclusion and involvement of a diverse scope of stakeholders in re-assuring that most, if not all, aspects of a complicated issue are brought to the attention of the policy-makers before policy choices begin to be considered.

Some argue that “policy analysts fail more often because they formulate the wrong problem than because they choose the wrong solution” (Dunn 1988, 720 in Vesely, 2001). This means that, problem definition takes priority over other phases of public policy analysis, such as formulation of variants and choosing among them. Moreover, defining complex social problems, such as undeclared work, requires a pluralistic approach since various stakeholders will invariably put forward diverging positions, and at times competing views. Still, because the policy choice, which follows from the understanding and formulation of the policy choice, is also a value choice, it is imperative that in a democratic society this value is congruent or at least agreeable with the dominant current in society (Dery, 1984). The way a policy problem is defined and framed also impacts on how the problem is owned and by whom, i.e. who has the perceived and/or enforceable authority on the way a particular social problem is discussed and acted upon in the public sphere (Rochefort and Cobb, 1995). Since undeclared work is a complex multi-actor social issue, it calls for open definition and ownership frameworks and for more attention to be paid to different rationalities and lines of argument. That is to say, narrow and authoritarian-style definitions and bureaucratic problem ownership, risks stakeholder alienation and ultimately the failure of policy.

According to Rochefort and Cobb (1995), problem definition “has to do with what we choose to identify as public issues and how we think about these concerns” (Rochefort and Cobb, 1995:8). Irwin (2003) maintains that “the task of problem definition requires a careful consideration of the parameters of an issue and the context within which a recommendation for a policy change will be made (in Vesely, 81). While Dery (1984) has stipulated that problem definition is “a never ending discourse with reality, to discover yet more facets, more dimensions of action, more opportunities for improvement (Dery, 1984: 6-7). He further proposes three ways of investigating problems.

*Problems as situations.* A problem is taken to mean simply a state of difficulty, or set of undesirable conditions. Therefore problems are objective entities and/or empirical phenomena, whereby the definition of the situation is often a description of the causes (Dery, 1984).

*Problems as discrepancies.* Problems are considered as gaps between what is and what should be. However, problems are defined as such only when a possible solution is conceivable or a correlating process is identified that is capable of bridging the gap (Dery, 1984) Viewing undeclared work through the prism of a discrepancy problem may be associated with a

divergence between planned state revenue and actual revenue, or expected compliance levels versus observed compliance levels. Often, this discrepancy is investigated by economists as market (e.g. labour market failure) and/or state failure which necessitates and justifies state intervention (see Dollery and Worthington, 1996). In recent years, however, undeclared work is being increasingly explained with the social context of institutional asymmetry, or the discrepancy between the formal and informal institutions in a society.

*Problems as opportunities* – Problems are perceived analytical constructs, therefore it is of essence to interrogate how a policy issue is defined. In this sense problems are not the same to all interested or even disinterested parties. Dery argues that problems are better treated as opportunities for improvement, since in such cases the process of problem definition will be “one of search, creation and initial examination of ideas for solutions” (Dery, 1984: 26). It is therefore paramount to secure feedback and generation of ideas from a wide circle of stakeholders, in order to reach a workable problem definition and further explore solutions feasibility.

Effective problem definition requires political will as, “unless a social condition has been consensually defined as a pressing social problem, and unless the condition has become fully politicized and debated, and the parameters of potential action agreed upon, there is little likelihood that policy-making bodies will be receptive to the results of social science research” (Weis, 1979: 427).

Often complex social problems are being framed for political purposes. The choice of problems is closely related to politics and many definitions are possible and each implicitly points to its remedy. According to Dery, problems are defined, and not “identified” or “discovered”. They cannot be detected as such “but are rather the products of imposing certain frames of reference on reality” (Dery, 1984: 4). Framing is prevalent in both media and political rhetoric. While framing makes it easier to understand and deal with complex problems it risks obfuscating the complete rationale and inner workings of an issue, therefore precluding problem redefinition and policy review. This may be particularly the case in organisations with poor institutional learning traditions and capacities. In very general terms framing is defined as “a way to meaningfully structure the social world” (Reese in Grim and Andsager, 2011). Entman (1993) elaborates further on what it means to frame:

*“Framing essentially involves selection and salience. To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described”*  
(Entman, 1993: 52)

In addition, it is argued that frames can define problems, diagnose causes, make moral judgements and prescribe remedies (Entman, 1993). Therefore, frames are employed in strategic communications of political and organisational actors (e.g., political parties), in the cognitive constructs of journalists, in news media content, and in the perceptions of the public. More importantly communicators, such as political elites, social movements, lobbyists, and activists develop frames about an issue and try to establish these in the public discourse and in the news

media (Matthes, 2012). This means that is that strategic actors, journalists, and the public do not simply digest a particular political and social reality. On the contrary, political and social issues are dynamic as they are being selected and interpreted in various patterns. The perceptions and understanding of issues are discussed, challenged, and revised over time. In this sense, frames are “selective views on issues—views that construct reality in a certain way leading to different evaluations and recommendations” (Mathes, 2012: 250). Complex social, economic and political problems are often difficult to be comprehensively conveyed and understood in their depth and scope particularly when they are a subject to public debate and political decision. Political and economic interests, however, also may cause skewness in interpretation.

Figuratively speaking, framing provides a ready-packaged piece of knowledge that explains a particular phenomenon. In the process of organisational learning, it may not be expected that all staff participate in defining and understanding a policy problem in the way that practitioners or academics do. Instead policy problems are often framed and accepted as truthful through a “black-box”<sup>4</sup> type process of knowing. Therefore, framing is crucial in the policy cycle as it underpins shortcomings of problem definition and policy response.

More often than not, the fight against the undeclared economy is framed in numbers, wherein its gravity is supported by the portion of GDP in the undeclared sector, while measures focus on direct controls setting numeric and quantitative intervention goals such as volumes of inspections, detections and sanctions<sup>5</sup>. Recent research, however, indicates that undeclared work should also be framed in terms of social values and public attitudes toward institutions, while evaluating interventions designed upon such framing seek to uncover whether and how attitudes have changed post-intervention. The framing of the undeclared economy in quantitative and numeric terms ultimately contributes to the design of policy approach that treats its symptoms and results, rather than its causes and drivers. In addition, as we will explore in the next part, the findings from evaluation studies of direct methods have no direct discernible link to the size of the undeclared economy.

In recent years developed states have been increasingly dedicating resources and attention to curative measures that foster commitment and aim at closing the gap of institutional incongruence, in what may be dubbed as a paradigm shift away from over-reliance on direct approaches, such as deterrence and prevention. This is the case since earlier understanding of undeclared work had been founded mainly on economic theory, explaining the decision to participate in non-compliant activities as a result of a rational cost-benefit analysis (Allingham and Sandmo, 1972, on rational theory see also Scott, 2000). This theoretical understanding of the undeclared economy empowered the predominance of the deterrence approach in policy design to tackle non-compliance. More recently however, arguments grounded in institutional theory and evidence from contemporary research have largely rejected the rational-actor models by placing personal and social norms and values at the centre of explaining and modelling predictions in non-compliance (Onu and Oats, 2014; Torgler, 2003, 2007; Wenzel, 2004). As

---

<sup>4</sup> A Black-box phenomenon is a type of system, where the initial stimuli and end results are known and observable, but the inner workings of the system remain unknown (see Bunge, 1963; Glanville, 2009)

<sup>5</sup> See, for example, Bulgarian PM Borisov’s remarks on the fight against the grey economy <http://www.europost.bg/article?id=14011>

such the undeclared economy is studied as a complex behavioural issue within a specific context. Further still, the interaction between the government and governed is examined in terms of the interplay of key factors such as individual and group perceptions of fairness, trust and justice. Thus, the relationships and dynamics within the context of a social contract between authorities and citizens are established as major determinants of compliance (Feld and Frey, 2002; Hartner et al., 2003; Murphy, 2003). Social norms and values have been employed in prediction models for tax compliance not only as the product of the relationship between state-individual, but also in terms of value formation within interpersonal interactions (Wenzel, 2004).

Consequently, we are faced with two important developments that define the social, cultural and economic context in the countries subject to the GREY research. First, that both undeclared work, which is regarded here as a “wicked problem”, and measures to tackle it, take place within a particular localised context, in which they interact (Williams, 2014). By defining undeclared work as a “wicked problem” we mean that there are “multiple drivers which display interdependencies and any attempt to tackle these drivers often has unforeseen and unintended effects” and that “there is no clear and correct solution that can be simply implemented which necessarily results in the desired outcome” (Williams, 2014:8). And second, that this context is subject to the interplay of the value-forming behaviours of state, policies and citizens. Put simply, both policies and citizens’ attitudes and beliefs can influence the context. In this paper we define institutional incongruence as the asymmetry between state institutions and rules and social values, norms and beliefs.

### ***Organisational learning and politicisation***

There might not be a universally accepted methodology for policy evaluation. That however may be beside the point. What matters, Rist argues, is whether “that which is done is found useful to those who need the information”, or whether policy evaluation has impacted upon the governance of society. What that means is perhaps the utility of evaluation is more important than its method. (Rist, xiii) In this there has been a transition in the way evaluation results reach their intended users: from delivering truths toward supplying data/information which is left at the discretion of the policy-maker as one of many factors in the decision-making process.

Evaluation research has been based on a rational paradigm, that has been based on three assumptions: 1. That policy goals are clearly defined; 2. Evaluation analyses the direct cause-effect relationship between policy means and policy goals; 3. Organisation guarantee that the policy is implemented in such a way that program activities are directed toward the achievement of its envisioned or planned goals (Aelbek, 1995: 11).

Chelimsky (1991) outlines three main “mismatches” in the utilisation of evaluation research that hint at the dangers of political and bureaucratic bias: (i) when political requirements are so overwhelming that information simply will not be sought; (ii) when information is sought, but contextual or resource constraints on the analysis impede researchers from actually producing the information needed for the decision; and (iii) when “state-of-the-art” research problems allow only inconclusive answers to decision-makers’ questions. In any of these situations, there are

constraints-either on the decision or on the research-that tend to prevent a match between the two from occurring (Chelimsky, 1991).

The problem with social science based policy evaluation research is that it can never completely be impartial and free of bias. As Aelbek (1995) puts it, “evaluation cannot offer objective truths and/or true knowledge, but rather good and relevant information” (Albaek, 1995:14-15). Therefore, as the very policies they try to assess, evaluations may fall prey to value and political bias.

The availability of information is crucial for a successful completion of the policy cycle. Politicization, however, may affect the availability of information, thereby changing patterns of organizational learning. Moreover, politicization may cause biases in the analysis of organizational failures. Time constraints are important factors in this respect and policy makers working under political agendas under pressure often have a restricted timeframes which may impede drawing lessons that make a link between policy failure and structural inadequacies in institutions. Instead, “public officials use rhetorical tactics or emphasize the role of individuals in a process of accountability and blaming, in order to depoliticize the problem in the short run” (Dekker and Hansen. 2014).

In this sense, frequent political sweeps and in-depth politicisation may cause bureaucratic amnesia and prevent good practices, such evidence-based policy evaluation, from taking hold in administrations. (Sander Dekker, Dan Hansen, 2004). This may also be the case since, policy officials in institutions may distort output and manipulate or spin data to favour or conceal policy performance (Bjørnholt and Larsen, 2014)

Deep politicisation may contribute to an “us against them” or clan mentality in administrations, therefore thinning public trust in the institutions’ ability and willingness to treat all citizens equally. It is yet to be thoroughly researched what the precise impact of politicisation on policy-making in South-East Europe is; however it may be assumed that politicisation plays an important role in the perception of administration’s ability to provide public services in an unbiased manner. Politicisation also inhibits administrations’ ability to initiate innovative policies and measures to tackle complex problems, such as undeclared work. For instance, a survey in the Bulgarian state administration concluded that:

*‘[c]ivil servants are in a constant fear of change, there is an absence of cohesion and team spirit and a lack of joint objectives. ... This generally leads to a situation where most civil servants prefer to take a ‘defensive position’ and attempts to perform their work by taking minimum risks thus avoiding responsibility for any conceivable error’ (Meyer-Sahling, 2004, p.83)*

As a consequence, and following a change of regime, second tier civil servants no longer have an incentive to be informative towards a successor party government. Instead, they tended to be passive in their policy making approach unless personnel policy is de-politicised (Meyer-Sahling, 2004). This environment of lack of incentives and motivation in the administration hinders the genesis of informed and innovative policies to tackle undeclared work and contributes instead to locking in the cycle of institutional incongruence regardless of which party

dominates government. In such circumstances, then, it is not uncommon for groups to look for a solution even before the problem is defined (Dery, 1984, p.57). Stakeholders in the fight against undeclared work, thus, may be inclined to search for solutions based on adapting or extending, and therefore, over-relying on policies that are familiar and tested, such as repressive or direct measures. Research, however, shows that the problem of undeclared work cannot be simply defined as non-compliance resultant from rational actor theory, (i.e. it is not a tame problem) but it is rather a complex, multi-contextual issue that includes, among other factors, trust in governmental institutions and public sense of judicial and administrative fairness.

Many of the policy processes are aligned with electoral cycles, therefore potentially intermixing policy with political results (Nunns, 2008). Since, on many occasions policy evaluation occurs at the end of the policy cycle, results from evaluation, if not the evaluation process itself, are at risk of being either disregarded or skipped entirely. A potential example might be a policy that has caused a public outcry or is extremely unacceptable within influential groups. The policy failure may be quickly hidden under a rug of political risk management with no consequent opportunity for evidence or evaluation based lesson-learning. Political actors are rarely concerned why policies work or not: they excel at taking credit for good policies and directing blame or scapegoating in instances of policy failure. In member states with monitoring mechanisms still in place, it may be argued that circumstances are even more complicated. In countries where there is a high level of politicised public administration, the electoral cycle necessarily changes a number of agency heads, while in certain organisations the political sweep may reach directorial levels. Meaningful policy evaluation in such scenarios is hard to imagine for several factors: 1. The persons/structures implementing and/or executing the policy may no longer be available; 2. Politicisation yields an anticipation of the next electoral political sweep by incumbent administrators, therefore public administration may be deliberately ripe with islands of knowledge and information ownership, rather than custodianship, factors making gathering of information for evaluation purposes extremely difficult.

### ***Capturing the context***

Social science knowledge does not readily lend itself to conversion into replicable technologies, either material or social (Weis, 1979). It is constructed in a dynamic and ever-changing contextualised complexity. Consequently, such complexity may not be handled by a single expert. Moreover “different political systems have developed different ‘civic epistemologies’ and ‘boundary arrangements’ in terms of how they decide the nature and role of knowledge and criteria for evidence, make sense of epistemic conflict and the mechanisms for resolving it (Pearce and Raman, 2014). In Bulgaria, Croatia and FYROM, it may be argued that this epistemic governance has been traditionally one of an authoritarian top-down approach with little or no stakeholder input, while policy problem definitions, agendas and evaluations have been ideologically loaded. Authoritarian regimes, in particular, “not only conceived of the world in certain recurrent, oversimplified categories, but also imposed its ways of seeing on people's lives”, therefore expanding the rift between formal and informal institutions and value systems (Jasanoff, 2004, p.26). This authoritarian intention was to “bind citizens' whole existence, not just their imaginations, to the service of the state's grid-like vision” (ibid). Furthermore, the

purpose of such grand plans of social engineering was to make citizens and their economic productions more "legible", that is, easier to count, survey, order, exploit and control.

Researchers may be prevented, again for reasons beyond their control, from giving decision-makers the information they need. The newness of a particular topic, for example, the immaturity of the research field, and the insufficiency of resources constitute contextual constraints on the work that may preclude the development of a useful research product (Chelimsky, 1991: 227). Wicked public policy problems, such as the informal economy represent complex social systems in which many diverse individual members and groups act independently and interact with each other. Such environments are adaptive in that individuals learn and adapt to new policies over time and their new behaviours converge to develop new and different collective patterns (Desouza and Lin, 2011).

It is important to note that policy-cycles are influenced by context and at the same time, are a factor which shapes that context. Therefore, an approach to evaluation is needed which recognises and captures this complex interplay of circumstances. A key requirement in the realist school of evaluation is to take heed of the different layers of social reality which make up and surround programmes (Pawson and Tilley, 2004). In this respect it is crucial to identify all stakeholders and possible drivers and motivations behind the irregular practices to be tackled through the proposed policy. The triggers of change in most interventions are ultimately located in the reasoning and resources of those touched by the programme. Effects are thus generally produced by and require the active engagement of individuals (Pawson and Tilley, 2004). Pawson and Tilley (2004) develop a framework or a way of looking at policies that captures their contextualisation in producing effective and accurate evaluation findings.

*Programmes are embedded.* As they are implemented, policies are embedded in social systems. It is through the workings of entire systems of social relationships that any changes in behaviours, events and social conditions are effected. Whether a policy is feasible will depend on analysing and investigating the four 'Is': *Individuals* – in their envisioned capacity within the proposed intervention, i.e. will they be able to comply with proposed administrative and procedural changes; *Interpersonal* features - relationships between policy-affected individuals; *Institutional* arrangements and characteristics – organisational capacity to implement policy; to conduct evaluation and engage in institutional learning; *Infra-structural* systems - the state of institutional asymmetry, are formal and informal values aligned; what is the dominant policy approach and policy-making style (Pawson and Tilley, 2004).

*Programs are active,* and as such, they only work through the stakeholders' reasoning. This implies that an understanding of the interpretations of programme participants and stakeholders is integral to evaluating its outcomes (Pawson and Tilley, 2004).

*Programmes are open systems.* Programmes cannot be fully isolated or kept constant, in that they are dynamic. Policies may change because of "unanticipated events, political change, personnel moves, physical and technological shifts, inter-programme and intra-programme interactions, practitioner learning, media coverage, organisational imperatives, performance management innovations and so on". Because such external factors always exert an effect on the

implementation of a policy, this requires that they are never quite delivered in the same way. Recognising that programs are dynamic, embedded in context and changing is to regard them as self-transformational, that is, they can change the very conditions that facilitated them in the first place (Pawson and Tilley, 2004).

In the realist approach, stakeholders are identified as key sources for producing programme theory and providing data on how the policy works. Stakeholders are recognised to generally have experience of and therefore expertise in specific stages and developments within an intervention. Realist evaluation requires data both on process and outcome, and on individuals, interrelationships, institutions and infra-structures. (Pawson and Tilley, 2004)

Identifying the crucial programme mechanisms, i.e. how exactly an intervention works, is only the first step in a realist evaluation. Realism stipulates that measure are only active under specific conditions, that is, in different contexts. Pawson and Tilley (2004) define context as “those features of the conditions in which programmes are introduced that are relevant to the operation the programme mechanisms. Realism utilises contextual thinking to address the issues of ‘for whom’ and ‘in what circumstances’ a programme will work. In the notion ‘context’ lies the realist solution to the panacea problem.” (Pawson and Tilley, 2004:7) Realism accepts that the policy success and failure are determined by whether the context is supportive of the initiative or not. Therefore, it is crucial to understand what programs are feasible within the context in which they are both designed and implemented.

The context in our countries of interest has recently been subject to increased scrutiny especially with regards to its role in shaping the characteristics of the drivers for engaging in non-compliant activities (Williams et al, 2014a, 2015a). However, it is becoming evident that this context must be considered in relation to policy-making and evaluation as well. Low tax morale, lack of trust in public institutions and persistent low levels of public perception of judicial and administrative fairness have been identified in Bulgaria, Croatia and Macedonia. This has generally led to the conclusion that the overall context in the three states is one of deep institutional incongruence, or put differently, a sharp divergence of formal and informal values (Williams et al, 2015b). Therefore, the prevailing context must be taken into consideration when defining policy areas, designing and evaluating policies.

### ***Information collection***

Designing effective policies for contextualised social problems presents a challenge for policy makers. Some regard the problem solving in the policy design as fundamentally information seeking through a vast maze of possibilities and source. Ideally, policy makers are meant to review and consider all possible solutions and find the optimal one. This is, however, rarely the case as policy makers are limited by their bounded rationality and computing capacity as human beings (Desouza and Lin, 2011).

Moreover, it is further suggested that different policy areas produce information with varied degrees of reliability, which may determine how policy makers respond to policy problems. Therefore it is expected that policy problems that are easy to quantify on an regular basis are associated with more evenly distributed political attention and more incremental policy

adjustments, while those that are only observed sporadically through less accessible means produce punctuated attention and policy responses, as problems unexpectedly surface (Boswell, 2012). In the case of the grey economy the collection and analysis of reliable information is crucial to policy-makers for designing sustained and adequate policies, particularly considering the difficult and often futile task of gathering intelligence on human behaviour whose main motivation is to remain hidden from the government radar. Therefore, in such context involving a wider circle of stakeholders is vital, with a view to collect the optimum available data to feed into and justify a potential policy cycle.

### *Timescales*

Research cycles typically take much longer than the policy cycles they aim to evaluate (Pawson, 2011, Nunns, 2009). There is an inherent mismatch in the timeframes associated with the policy process and evaluation activity. Policy processes are aligned with the electoral cycles, with policy making and funding for new initiatives typically occurring at the beginning of the mandate period. If an evaluation is scheduled towards the end of the policy cycle, its findings will be too late to be used for decision-making purposes (Nunns, 2009) Therefore, the role of an ex ante feasibility assessment involving stakeholders early on in the policy design stage may prove time and resource saving exercise, particularly scarce in times of economic crises and/or austerity measures.

#### 4.1. Evaluating measures to tackle the grey economy – some lessons

Working papers No 1-3 provided the baseline assessments on the nature and scope of the grey economy in Bulgaria, FYROM and Croatia, as well as evaluation of the overall policy response in these states. The challenges in assessing policies and measures were associated with a general lack of consistent evaluation practices in all three states, and where such have been performed, data had been scarce on methods, methodologies and tools used in the evaluation studies (Dzhekova and Williams, 2014). In Bulgaria, research established the following deficiencies:

*“overall, the national policy approach towards the undeclared economy is characterised by the following deficiencies: 1) lack of systematic evaluations and cost-benefit analysis of measures; 2) insufficient tailoring of measures to the features of the national economy, as well as fitting those into the existing policy infrastructure; and 3) economic and administrative measures neutralise their respective effects” (Dzhekova and Williams, 2014:8)*

In all three countries, the **direct policy approach**, i.e. control measures on detections, sanctions and prevention, had been found to be the dominant one in tackling the undeclared economy. Consequently, most of the available policy evaluation results had been mainly produced by top-down approaches focusing exclusively on **positivist quantitative methods** (for example increased and/or joint inspections initiatives and campaigns in Bulgaria and Croatia). This coincides with findings demonstrating that the dominant approach in evaluation of EU funds at the European and national level is one of positivist quantitative methods, while others are rarely used (Hoerner and Stephenson, 2013). Keeping in mind that evaluation practices and routines in Bulgaria, Croatia and FYROM, have been largely externally imposed, further research may be

needed in exploring the potential impact of EU requirements and methodological guidance for policy assessment on the overall national approaches and styles of policy evaluation research in these countries.

Lack of policy evaluation, particularly at the consultative and feasibility stages in all three states is endemic. In Bulgaria, despite ex ante impact assessments being legislated as requirements in a number of laws (for example the Act on Limiting Administrative Regulation and Administrative Control on Economic Activity), studies and investigations have established that such are not being conducted (World Bank, 2010). In Croatia, the majority of interventions were conducted without pilot initiatives or public debates, and there is a general lack of studies evaluating the effectiveness of the introduced strategies (Francic and Williams, 2014). In FYROM, government stakeholders in the fight against undeclared work, such as the Employment Service Agency for example, are still without a comprehensive approach to monitoring and evaluation, while preparations in this area are not very advanced (European Commission, 2011).

Given that proper evaluation research is often absent, it remains the futile task of the researcher to pinpoint the probable causes for policy failure. Where policies have failed, it is often stated how they failed, such as that overly complicated procedures made compliance an undesirable option, but no explanation is provided as to why these specific results had been reached. An analysis of policy failures provided in the preceding working papers shows similarity in the probable causes of policy failure in the three states, particularly in Bulgaria and Croatia. Most of the identified intervention failures point to a logic that links them to inadequate or non-existent feasibility evaluation of proposed and implemented measures. Or in other words, policy failure may have been avoided if a wider circle of stakeholders had been involved in a feasibility assessment of the policy initiative at an early stage, starting with the policy issue definition (see Table 2). Where stakeholders are, in fact involved, it is a common practice that institutions do so **after** the policy measure has been already designed. In Bulgaria, for example, non-governmental organisations are increasing voicing concerns that local and central authorities are consulting with them on proposed policies pro forma and post factum, leaving few opportunities for policy problem re-definition and intervention adjustment (Dnevnik, 2015; LIL, 2015). Moreover, despite a number of EU-wide recommendations and guidelines issued by the Commission highlighting the importance of stakeholder consultation in policy-making, a recent report on consultative practices in Bulgaria uncovered that most of the 62 existing consultative bodies do not function effectively (Bulgarian School of Politics, 2014). Some of the deficiencies identified in the report are: a lack of feedback from state institutions; exploiting the consultative bodies in order to legitimise an already designed measure; lack of clear rules, procedures and communication policies between state institutions and stakeholders (Bulgarian School of Politics, 2014).

The lack of genuine stakeholder involvement with political support in the definition of policy problems and feasibility evaluations is identified in a number of policy failures in Bulgaria and Croatia. In many cases, both stakeholders within and outside government had not been adequately involved in feasibility studies. A prominent example is the introduction of Ordinance 3 by the Bulgarian NRA requiring all oil trading, processing and producing companies to install

measuring and monitoring devices. The measure was met with staunch resistance from business as it was argued that the cost of compliance is too high. More importantly, however, conflict stirred between the Revenue and Customs agencies, as the latter were tasked with enforcing and overseeing the ordinance. Customs resisted the new measure as they maintained they lacked the necessary resources for enforcement (Dzhekova and Williams, 2014). In Croatia, it is argued that the most important weakness of the fight against undeclared work are the too frequent and instant changes of legislation which do not leave enough time for adaptation, not only for the target population, but also for the authorities. One such example happened in the beginning of 2014 when inspectors were left without clear guidance on reallocation and responsibilities after the abolition of the State Inspectorate (Francic and Williams, 2014:41). Two more measures in Croatia failed to achieve desired objectives because the lack of stakeholder involvement in feasibility assessments precluded policy-makers from acquiring actual knowledge of the circumstances, attitudes, abilities and needs of the policy targets and beneficiaries. The Act on Job Retention, which came into force in 2009, failed to fulfil its purpose due to over-restrictive eligibility criteria, which resulted in only few employers receiving the subsidy (Francic and Williams, 2014). The attempt to regulate nannies in Croatia also proved inefficient due to restrictions imposed on eligibility, accompanied by complicated procedures and unclearly defined responsibilities of governmental bodies in the registration process (Francic and Williams, 2014). Such developments point to non-existent or inadequate feasibility evaluations by failing to recognise the circumstances, needs and attitudes of policy stakeholders.

Politicisation, political arithmetic and bias, and bureaucratic logic may be identified as the main culprits for policy failure in a number of other measures. For example, stakeholder consultation in FYROM has been plagued by marginalisation and politicisation of trade unions, with the latter progressively losing credibility among their members and influence with employers and the government. For instance, after the adoption of the collective agreement in 2009, the government pushed through some amendments to the Law on Labour Relations regarding the activity of labour unions, which affected their rights negatively. This was done without holding the required social dialogue consultations (Dzhekova et al., 2014: 43). In Croatia, an ambitious award-winning initiative to improve the overall administrative functioning of the government, “HITRORES”, failed because the state rejected the majority of proposed reforms and recommendations.

**Table 2. Policy failures and probable causes**

Measure	Policy Failure	Probable causes
<b>Act on Limiting Administrative Regulation and Administrative Control on Economic Activity (Bulgaria)</b>	Lack of awareness campaign	<ul style="list-style-type: none"> <li>• Inadequate/flawed policy design – no awareness campaign envisioned.</li> </ul>
<b>NRA Ordinance 3</b>	Stakeholder/partner failure to implement	<ul style="list-style-type: none"> <li>• Lack of stakeholder consultation and agreement.</li> <li>• Insufficient knowledge of policy target groups, beneficiaries and stakeholders.</li> <li>• Lack of stakeholder involvement in feasibility evaluation of the measure.</li> </ul>

<b>Structural changes in the Labour Inspectorate and The Act and Tourism Inspection Act (Croatia)</b>	Insufficient resources and provisions for control activities.	<ul style="list-style-type: none"> <li>• Inadequate involvement of stakeholders in policy feasibility study left inspectors without resources and guidelines.</li> <li>• Inadequate policy impact assessment.</li> </ul>
<b>HITROREZ (Croatia)</b>	Project recommendations not taken up by government	<ul style="list-style-type: none"> <li>• Bureaucratic logic</li> <li>• Politicisation</li> <li>• Political arithmetic</li> </ul>
<b>Act on Job Retention Subsidies (Croatia)</b>	over-restrictive eligibility criteria	<ul style="list-style-type: none"> <li>• Lack of or ineffective involvement in stakeholders in feasibility evaluation.</li> <li>• Inadequate knowledge of policy targets/beneficiaries needs, compliance abilities, circumstances and attitudes.</li> </ul>
<b>Registration of nannies (Croatia)</b>	over-restrictive eligibility criteria, over-complicated procedures	<ul style="list-style-type: none"> <li>• Lack of or ineffective involvement in stakeholders in feasibility evaluation.</li> <li>• Inadequate knowledge of policy targets/beneficiaries' needs, abilities circumstances and attitudes.</li> <li>• Inadequate policy impact assessment.</li> </ul>

It must be noted that many of the typologies of policy failure observed in Bulgaria, Croatia and FYROM, are identified in other EU countries as well. For example, an attempt to regulate nannies in Poland produced very similar results as in Croatia and failed for the same reasons (Eurofound, 2009a). In the Czech Republic a measure to increase employment rates proved controversial as it was designed with little concern for the actual needs, attitudes and abilities of the policy stakeholders (Eurofound. 2013a). The measure had been described as demeaning for those who participated, because of unrealistically strict control and registration requirements.

The Eurofound database provides a useful starting point for identifying some of the practical and actual trends and challenges of evaluating measures to tackle the undeclared economy across European states. It should be noted, however, that basing an analysis on the evaluation feedback provided in the database by practitioners and experts may have serious drawbacks. Firstly, not all measures have been evaluated, as many of them are still in their implementation phase, therefore narrowing the selection pool to measures prone to speedier and simpler evaluation methods. As such, the analysed evaluation feedback may not be considered as a fully representative sample. Secondly, the entries do not provide a comprehensive illumination into the methods and methodologies used to produce the evidence of the evaluation. Therefore, the summary and analysis presented here is to be viewed with caution and considered as a tentative exercise at risk of examining the subject matter superfluously until a more broad research is attempted.

From an evaluation perspective, repressive measures could be a preferred choice since simple quantitative research might be able to provide sufficient data to policy makers – increase/decrease in detections, increase/decrease of sanctioned persons/entities, increase/decrease in collected penalties and so on. An evidence-based realist approach would take into consideration the diversity of context surrounding the said measure. Repressive

measures produce numbers, and numbers are a weak spot for proponents of the positivist approach or those over-reliant on quantitative method proclaiming the “numbers don’t lie” motto in public policy. Although research is fast pointing to the need and potential promise of indirect measures, many states who have implemented such instruments do not see them as effective (Dekkar et al, 2010). It may be argued that one reason for such a perception is the fact that evaluation of behaviour changing measures, such as awareness campaigns, may only be conclusive when performed periodically after the measures has been completed in order to determine change in behaviour over time. For example, evaluation feedback of awareness campaigns in Montenegro, Estonia and Iceland identify the difficulties in adequately measuring and linking participant behaviour change, objectives of campaign and overall effect on the undeclared economy (Eurofound, 2009b, Eurofound, 2013b, 2013c, 2013e). This is also the case with more comprehensive qualitative evaluations aiming to assess long term policy effects. For instance, in an unprecedented effort, the Bulgarian Ministry of Labour is currently qualitatively evaluating the effect of policy spending to increase employment rates in the period between 2003 and 2011. Four years after the end of the evaluation target period, the assessment study is yet to be finalised with only partial evaluation results being published (Koleva, 2015). In addition, contingent on policy and evaluation timeframes, it will not be uncommon that evaluation results of an awareness or commitment measures are produced after election cycles and/or administrative change, therefore potentially exposing evaluation results to political bias and arithmetic. Yet, from an evaluation point of view, the discussion is in its infancy, as further research is needed to examine how indirect policy measures are being evaluated, particularly with regards to their impact on the size and prevalence of the undeclared economy. The most important question is whether their evaluation is contextualised according to the dynamic realities of the domains that indirect measures are trying to meddle with – such as tax morality, or are evaluation methods used to assess repressive measure simply extrapolated with little adaptation. At any rate, it must be further mined into the specific reasons why certain indirect policy measures are being deemed as inefficient by many policy makers and stakeholders. However superfluous this observation may be at this juncture, it serves to point out a gaping void in evaluation research of policies tackling undeclared work across the EU. It is paramount to investigate the reasons, rationales and methods of evaluation behind negative and positive assessments of policy-makers and stakeholders. After all, what we are ultimately concerned with here is not professional opinion but professional evaluation.

The feedback on evaluations of measures to tackle the undeclared economy provided in the Eurofound database generally reflects some of the theoretical considerations and risks outlined in the previous parts. The database of evaluation feedback proves quite useful in constructing an initial glance of practical evaluation issues as findings may be organised per type of measure utilising the framework developed by Williams and Renooy (2009).

**Table 3. Policy approaches and evaluation risks**

Approach	Method	Evaluation Risks, Challenges and Dilemmas
<b>Deterrence (pursue and punish)</b>	<u>Improving detection</u> Increasing penalties Increasing perception of risk	<ul style="list-style-type: none"> <li>Evaluations tend to be “dehumanised” and focus on positivist quantitative methods. No immediate discernible connection may be evidenced between number of inspections, detections and sanctions, and the size and prevalence of the undeclared economy.</li> <li>Levied sanctions and collected fines, combined with increased number of control activities are often cited as evidence of a successful policy.</li> <li>Traditional cost-benefit and cost-efficiency evaluation do not illuminate on the actual impact on the undeclared economy.</li> </ul>
<b>Enabling formalisation/ compliance</b>	<u>Prevention (deter entry)</u> <u>Curative/stimulating (encouraging movement out of shadow economy)</u> Fostering commitment	<ul style="list-style-type: none"> <li>New regulations (e.g. licensing and registration regimes) often remain obscure on account of poor stakeholder involvement and/or inadequate awareness campaigning, therefore pre-determining poor evaluation results.</li> <li>Qualitative methods, such as surveys, are often only budgeted to capture a snapshot evaluation of commitment and curative measures, which proves inconclusive as measures are designed for long term effect of changing social behaviour. Therefore, the risk of misaligned timeframes of evaluation and program cycle is significant in such measures.</li> <li>Policy packages, combining several types of measures and including input across agencies, require the collection and analysis of evaluation data from multiple sources via multiple methods. It is often unclear how to measure the combined output and results of such packaged measures with regard to their impact on undeclared work.</li> </ul>

*Source: derived from Williams and Renooy (2009)*

The Eurofound database presents a unique opportunity for collaboration among policy officials, evaluators and academia across participating European states. This database could be easily expanded to include more interactive feedback and discussion functions for each policy measures and/or group of policy measures, whereby experts could exchange ideas on their feasibility in different contexts and make recommendations on issues related to policy evaluation.

## 5. Toward a genuine collaborative approach – bridging the gap of institutional asymmetry

One way of looking at the role of context and how it impacts on the development of evidence-based policy-making is to consider political, cultural and social orders as “being produced together” where “knowledge, and its material embodiments are at once products of social work and constitutive of forms of social life” (Jasanoff, 2004). Indeed, policy developments, particularly in the democratic world, are increasingly indicative of and responsive to a dynamic change in context where morals, values and their legal representations intermingle. Legalisation of gay marriage, marijuana and euthanasia in many Western jurisdictions are just the more prominent examples of innovative-type policy measures driven by an ever-changing context. What is more important however, in the above examples, is that they represent complex social problems requiring adjustment in individual and group behaviour. Even more crucially, a wide circle of stakeholders have been involved in brainstorming, testing and designing policy responses, particularly where decisions had been taken outside voter discretion.

In addition, policy innovation appears to be enabled by a certain level of mutual trust between citizens and their governments, among other factors. For example, context and trust have come recently to support an innovative stakeholder-driven policy approach to tackling employment and undeclared work in the Dutch city of Utrecht. The Netherlands has the highest proportion of part-time workers in Europe at around 46%, while at the same only 1.5% abuse the welfare system. The employment context and the low level of welfare-abuse create favourable conditions for a trust-based social policy experiment. The experiment, designed in cooperation with the University College of Utrecht, will test if the concept of *basic income* - a universal, unconditional form of payment to individuals, which covers their living costs – will be an effective approach to improving conditions for beneficiaries of the social welfare system (Dore, 2015). Regardless of the outcome of this experimentation, such recent developments underline the importance of trust-enabled stakeholder collaboration in producing innovative policy solutions.

Indeed governments in developed democracies have been increasingly coming to terms that complex social problems cannot be solved by a top-down strategy and a more collaborative approach is needed (APSC, 2007a, 2007b; House of Lords, 2013). Furthermore, stakeholder-driven approaches to policy evaluation have been established to possess a distinct merit, particularly in circumstance where traditional or conventional assessment methods fail to illuminate the complete aftermath of an intervention (Palenberg, 2011). The Magenta book, for example advises that

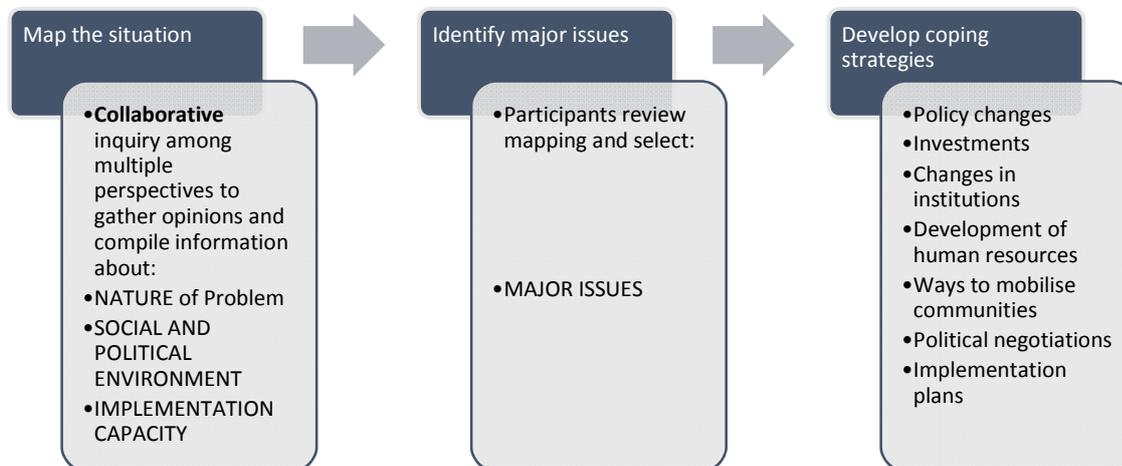
“inclusion of recipients, delivery bodies or stakeholders – through a steering group, for example – enhances the potential learning from an evaluation and acceptance of its results, but it has to be actively managed as a continuous process of communication and engagement” (HM Treasury, 2011a: 23).

Stakeholder participation, active input and feedback enhance overall awareness of the evaluation; improves feedback on research design and ensure effective communication of feasibility, interim and final conclusions. From an accountability perspective, transparency must be a feature of any evaluation but especially for a high-risk or innovative policy intervention. Transparent evaluation planning can determine the evaluation goals, questions and indicators, how the process itself will be brought to fruition, as well how evaluation results will be interpreted. This facilitates stakeholder engagement, allows the issues and risks to be identified and managed, and the delivery outputs and milestones to be agreed and documented. Also, evaluation research should be disseminated with adequate methodological detail for stakeholders to assess for themselves the robustness of the results (HM Treasury, 2011a).

To secure a meaningful stakeholder involvement, collaborative strategies must be developed. Collaborative problem solving is considered as being the most effective in dealing with wicked problems that have many stakeholders amongst whom power is dispersed (APSC, 2007a, 2007b). It is particularly relevant where part of the solution to the problem involves sustained behavioural change by many stakeholders and/or citizens, such as is the case with undeclared work. At the core of collaboration is a win-win view of problem-solving. Partnerships, joint ventures, whole of (or joined up) government, international treaties and information campaigns to influence lifestyle choices are all variations on this strategy (APSC, 2007a,b). The key advantages to employing collaborative stakeholder strategies include higher stakeholder commitment, more comprehensive and effective solutions, and fewer resources having to be used by any one stakeholder.

The handling of wicked problems, such as the undeclared economy, requires holistic rather than linear thinking (APSC, 2007a, 2007b; Williams, 2014, European Commission, 2007). This is thinking capable of grasping the big picture, including the interrelationships between the full range of causal factors and policy objectives. By their substance complex social problems are inadequately understood, and so initial policy designs that are drawn too narrowly, such as top-down positivist approaches, may lead to a mistreatment of what is essential in tackling the wicked issues. Linear and top-down approaches to problem definition and policy design are considered inadequate in generating innovative and effective policy measure to address complex social problems.

Figure 1. Collaborative policy consultation model (White, 2002: 164)



In considering a bottom-down collaborative approach to policy evaluation in Bulgaria, Croatia and FYROM we must first identify what context and important factors may necessitate the adoption of such an approach.

**Wider context:** As previously noted the predominant context of all three states is characterised by a state of institutional asymmetry (i.e. a persistent divergence between formal and informal values), low tax morale and low public trust in formal institutions, wherein stakeholders and the general public may not “buy-in” the presented policy proposals and evaluation results (Williams, 2014; Williams, 2015; Williams and Horodnic, 2015)

**Policy Evaluation:** As part of the encompassing policy cycle, programme evaluation is considered a policy mechanism in itself. Therefore it is indicative of the relationship and attitude between government and its citizens and as such it possesses a potential to carry a value-forming effect between citizens and their government (Lascoumes and Le Gales, 2007).

**Institutions:** Politicised institutions, observed in counties in SE Europe, have difficulties producing innovative solutions to complex social problems and are sluggish in achieving organisational learning and development, therefore lacking, at times, both motivation and capacity to tackle problems, such as the grey economy, on their own (Dekker and Hansen, 2004; Grzymaña-Busse, 2003; Meyer-Sahling and Veen, 2012; Meyer-Sahling, 2004).

**Problem definition:** Recent research points to social behavioural factors as major drivers of non-compliant behaviour (cite). Tax morale levels have been linked with a propensity to engage in undeclared economic activities, indicating the gravity of the gap of institutional asymmetry (Williams and Franic, 2015; Williams et al, 2015). Consequently, undeclared work has been

largely recognised as a wicked problem (Williams, 2014). However, because of the linkage between public trust and tax morale to rates of engagement in the informal economy, we also define the problem of undeclared work as one of opportunity to enhance institutional trust, foster collaboration and innovation, improve public policy effectiveness, and move toward bridging the gap of institutional asymmetry (Dery, 1984).

Stakeholder collaboration in the initial stages of policy evaluation or feasibility assessment appears as a best suited and logical step in the right direction within our context for a number of reasons. Firstly, the policy landscape in the EU on tackling the undeclared economy is significantly rich and diverse with hundreds of measures being implemented nationally and regionally (see Eurofound database). Policy transfer has intensified in recent years as local governments are tempted to look for quick-win policy-making by adopting and adapting successful measures implemented elsewhere into their particular reality. Therefore, there is a considerable risk to overlook stakeholder identification and consultation, as the “borrowed” measure may be deemed as successful by default, therefore omitting key players and factors in the evaluation process. Secondly, ex post evaluations of policies are not the only way of establishing what works. There are other approaches to producing such evidence that can be as informative, and may be less resource intensive and more timely. Pilot studies, focus groups or social experiments (see Slemrod et al, 2001; Polak et al. 2015), testing out a policy for a limited time or on a limited population, may be appropriate tools to learn about policy impacts where outcomes are uncertain. Having more regard to theories of human behaviour is also advisable. In short, even the question of what works can be addressed by a wider repertoire of evidence-gathering research (Solesberry, 2001). Thirdly, social interventions into complex social problems, such as undeclared work, occur in a contextualised setting that is formulated by the interplay between policies and problem stakeholders. Given this dynamic complexity, social interventions/policies are often subjective and unpredictable, making stakeholder involvement key to understanding what works, for whom and within what context. Finally, opening the early stages of policy design to a wider circle of stakeholders is aligned with the EU’s goals for improved governance and policy-making. In its strategy for enhanced democratic governance the EC has called for “a stronger interaction with regional and local governments and civil society” through a systematic dialogue with stakeholder at an early stage of policy cycles (European Commission, 2001).

In our previous analyses, we identified the lack of stakeholder involvement in the policy feasibility evaluation stage as a probable and most likely cause of a number of policy failures in Bulgaria, Croatia and FYROM. Stakeholder collaboration then appears as a logical safeguard against policy failure and a facilitator for the formulation of more effective and innovative measures. The benefits of stakeholder collaboration strategies in policy feasibility evaluation are manifold.

***Epistemic governance*** – The idea of epistemic governance suggests that the production of knowledge for governance itself needs be governed. But such knowledge is insufficient if it is abstracted from the political side of policymaking in which problems are framed and knowledge given meaning (Pearce and Raman, 2014: 388). What this implies is that problem definition is

not be usurped by political elites or administrations and should possess a stakeholder-driven dimension to ensure that epistemic governance is not irreversibly contaminated by political arithmetic and bias.

***Trust building, transparency and accountability*** – It is often argued that the most critical time frame for organizational participants to develop trust is at the beginning of the relationship (Mcknight et al, 1998). Therefore early stakeholder involvement in the policy feasibility evaluation process fosters engaged accountability, which is actionable on account of difference in expert opinion and/or new information brought to problem definition. Furthermore, stakeholder buy-in into a proposed policy implementation ensures ownership of initiative and therefore enhanced commitment to proposed problem solution. Broadening the network of the initial stages of problem definition and policy design also diminishes the possibility of bias as policy evaluation methods and considerations are shared among stakeholders.

***Innovation capacity*** – It is generally agreed that innovative ideas are initially incomplete and require further insight before yielding a valuable and practicable innovation. Finding the complementary piece of the idea requires that it is disseminated and shared. Circulation of ideas ensures efficient completion (Hellman and Perotti, 2011). Incomplete ideas need to be circulated and to be combined with insight by others before a valuable novel factor combination emerges. It must be noted however, that stakeholders must perceive the long-term benefits and continued incentives to participate, as disenfranchisement from collaborative work is likely to occur if initial incentives and commitment by leaders is withdrawn prematurely. (Bruttel and Friehe, 2014)

***Cost reduction*** – Early stage stakeholder-centric policy evaluations help safeguard against policy failures and mid-stage policy adjustments which are usually resource intensive (Dzhekova and Willaims, 2014). Moreover, ex post evaluations tend to be lengthy and expensive (Solesberry, 2001; Nunns, 2009)

***Information gathering*** – As previously noted, the manner in which information is collected and the manner and frequency its analysis reaches policy makers has an impact on the decision-making process (Boswell, 2012). Because activities in the undeclared economy are by default hidden from the government radar, there is a risk for policy-makers of not being adequately informed about a particular issue which may result in poor problem definition and ineffective intervention. Bringing stakeholders on board in the fight against undeclared work may provide indispensable problem-related information that fills in the knowledge gap.

Given that we have chosen to consider the problem of the undeclared economy as a wicked problem, but nevertheless one where there is opportunity for improvement, we must consider the practicability of reaching a win/win approach of feasibility assessment of a policy proposal. A win-win may be defined as an approach exceeding the best initial expectations (Nagel, 2002:5). In our case, the initial expectation of genuine participation of a wider group of stakeholder is to produce effective policy design. The added value in the win-win sense is that employing accountable, transparent, pluralistic and collaborative methods in the initial stages of policy-making will produce a trust building effect, increase the sense of policy ownership among

stakeholders, thereby potentially increasing compliance by having a diminishing effect on the state of institutional asymmetry. Therefore, we consider a stakeholder collaborative strategy to problem definition and policy feasibility assessment as a win/win model for policy-making.

## 6. Conclusion

In this paper, we have set out to delve into the complicated field of policy evaluation and the challenges associated with evaluating measures to tackle the undeclared economy. We took an exploratory approach in reviewing relevant theory and available practical experiences and feedback in order to identify challenges and shortcomings in policy evaluation. This exploration was set intentionally in the context of the knowledge and evidence produced at the earlier stages in the GREY project and published in the preceding working papers. This research had shown that undeclared work and the policy realm intended to handle it, occur within and are contingent on a complex social context within which they interact and impact on one another. We built out presentation and argumentation on evidence that this context in Bulgaria, Croatia and FYROM is characterised by a persistent state of institutional asymmetry. Furthermore, drivers arising from this context, such as low tax morale, are evidenced to explain in part the propensity to engage in activities related to the undeclared economy in all three states. In addition, the policy-making process had been found to suffer from a debilitating authoritarian legacy of top-down positivist problem definition and policy evaluation styles that fail to capture the full scope and vital peculiarities of the undeclared economy. Policy evaluation has been a relatively new administrative phenomenon in the GREY countries, as it has largely been externally imposed by EU-funding regulations and although it has been legislated into a requirement in local laws and transposed EU-law, evaluations are not effectively embedded in the policy-making institutional psyche.

Our preliminary analysis of a limited sample of available policy evaluations and results demonstrates that a common probable cause of policy failure with regards to measures to tackle undeclared work in the GREY countries is an apparent lack of knowledge and understanding of the needs, attitudes and abilities of policy beneficiaries and targets. This strongly suggests non-existent or inadequate engagement of all stakeholders in the initial policy evaluation or feasibility study stage of the policy-making process. Therefore, a stakeholder collaboration strategy is proposed as a remedy and safeguard to ensuring that policy feasibility evaluation at the onset of policy design is provided with sufficient relevant information about policy issue definition and delineation, beneficiary needs, attitudes and abilities, as they relate to envisioned policy outputs. Enhanced knowledge and adequate policy-relevant evidence are paramount to producing holistic, effective and innovative solutions to wicked social problems.

This exploratory effort has identified the collaborative stakeholder strategy to tackling the undeclared economy as a major recommendation for moving forward. However, diverging local and national economic contexts prevent us from formulating a stricter framework. The nature and scope of undeclared work could require different approaches and formulas to engaging stakeholders, e.g. differences may arise in the prevalence of seasonal work (tourist and agricultural), prevalence of legal and illegal economic immigrants, size and national importance

of worst afflicted economic sectors etc.) As such, any structured methods must be developed in consultation with the stakeholders themselves. Nevertheless, several minimum requirements must be observed:

- A **stakeholder identification** process must be completed. This should include intra-governmental actors, business, NGOs and academia. Policy-responsible officials may perform initial stakeholder scanning. A second round of consultative stakeholder scanning may be appropriate, among the initially selected ones, to ensure that all interested parties are invited in the policy discussions.
- Stakeholder collaboration is only meaningful **before** the problem definition and policy design occur. Such collaboration must be inspired by engaged or pro-active public accountability, whereby participants may advise and revise policy proposals.

More critically, however, adopting and institutionalising a stakeholder collaboration strategy to tackle complex social problems, such as the undeclared economy, constitutes a win-win solution when placed in the wider context of the GREY countries. Stakeholder collaboration is seen as a tool for building and improving institutional trust and ensuring stakeholder commitment to proposed interventions. Stemming from institutional theory we accept that policies, together with their tools and instruments reflect the government's attitude toward the governed, thus having an effect on the trust-building relationship between state and citizenry (Lascoumes and Le Gales, 2007). This means that policy instruments, such as policy feasibility evaluation, play a vital role in the trust-creation process between the state and its citizens. Based on this, we recommend that policies, as well as their tools and instruments are designed with built-in consideration for fostering trust, commitment and voluntary compliance. Such asymmetry diminishing measures may include increased reliance on collaborative approaches, externalising ex ante policy evaluations, extending power-sharing in the decision-making process to a wider circle of stakeholders, among others. The efficient inclusion of the incongruence-diminishing dimension to policy-making, however, must be preceded by an *institutionalised belief* in the attitude-forming and trust-building powers of policy measures, tools and mechanisms.

## Bibliography

Albaek, E., Policy Evaluation: Design and Utilisation. (1995). In Policy Evaluation: Linking Theory to Practice, Ray Rist (ed.). Aldershot: Edward Elgar.

Allingham M.G., Sandmo A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*. 1(2): 323–338.

Armstrong, H. and P. Wells. 2006. ‘Structural Funds and the Evaluation of Community Economic Development Initiatives in the UK: A Critical Perspective’, *Regional Studies*, 40, 2, 259–72

Australian Public Service Commission (APSC). (2007a). *Changing Behaviour - A Public Policy Perspective*. Sydney: Australian Public Service Commission.

Australian Public Service Commission (APSC). (2007b). *Tackling Wicked Problems - A Public Policy Perspective*. Sydney: Australian Public Service Commission.

Bachtler, J. and C. Wren. 2006. ‘Evaluation of European Union Cohesion Policy: Research Questions and Policy Challenges’, *Regional Studies*, 40, 2, 143–53.

Bjørnholt, B & Larsen, F (2014).The Politics of Performance Measurement: 'Evaluation use as mediator for politics" *Evaluation*, 20(4), pp. 400-411.

Boswell, C. (2012). How Information Scarcity Influences the Policy Agenda: Evidence from U.K. Immigration Policy. *Governance: An International Journal of Policy, Administration, and Institutions*, 25(3), 367–389

Bruttel, L and Friehe, T. (2014). Can short-term incentives induce long-lasting cooperation? Results from a public-goods experiment. *Journal of Behavioral and Experimental Economics*, 53, pp.120-130.

Bulgarian School of Politics “Dimtry Panitza” (2014). *A Study of the existing consultative councils and internet forms of citizen participation*. Bulgarian School of Politics “Dimtry Panitza. Sofia.

Bunge, M. (1963). A general black-box theory. *Philosophy of Science*, 30 (4), 346-358.

Cabinet Office. (1999). *Modernising government*. The Stationery Office Limited: London.

Cabinet Office (2004). *Personal Responsibility and Changing Behaviour: the state of knowledge and its implications for public policy*. The Stationery Office Limited: London.

- Contandriopoulos, D. and Brousselle A. (2012). Evaluation models and evaluation use. *Evaluation*, 18(1) 61–77.
- D’agostino, F. (1996) Social science as a social institution: neutrality and the politics of social research. *Philosophy of the Social Sciences*, 25(3), 396-405.
- Davies H., Nutley S., and Smith P. (2000). Introducing Evidence-based Policy and Practice in Public Service. In Davies H., Nutley S., and Smith P., (eds) *What Works: Evidence-based policy and practice in public service*. The Policy Press. Bristol.
- Dekker H., Oranje E., Renooy P., Rosing F. and Williams C.C. (2010). *Joining up the fight against undeclared work in the European Union*. Brussels: DG Employment, Social Affairs and Equal Opportunities.
- Dekker S., Hansen, D. (2004) Learning under Pressure: The Effects of Politicization on Organizational Learning in Public Bureaucracies. *Journal of Public Administration Research and Theory*, Vol. 14, no. 2, pp. 211–230.
- Desouza, K., and Lin, Y. (2011). Towards Evidence-Driven Policy Design: Complex Adaptive Systems and Computational Modeling. *The Innovation Journal: The Public Sector Innovation Journal*, 16(1), article 7.
- Dery, D. (1984). *Problem Definition in Policy Analysis*. Kansas: University Press of Kansas.
- Dnevnik Online (2015, 6 July). *Неправителствени организации към Фандъкова: Искаме нов начин на обществени обсъждания* [Non-governmental organisations to Fandakova: We want a new way for community consultations]. Retrieved from <http://www.dnevnik.bg>.
- Dollery, B., and Worthington, A. (1996) The Evaluation of Public Policy: Normative Economic Theories of Government Failure. *Journal of Interdisciplinary Economics* 7(1):pp. 27-39.
- Dore, L. (2015, 25 June). Dutch city of Utrecht to experiment with a universal, unconditional 'basic income'. *The Independent Online*. Retrieved from <http://www.independent.co.uk>.
- Dzhekova R., Williams C.C. (2014) *Tackling Undeclared Work in Bulgaria: A Baseline Report*, GREY Working Paper no. 1. Sheffield: Sheffield University Management School. 2014.
- Dzhekova, R., Franic, J., Mishkov, L., and Williams, C.C. (2014). *Tackling the Undeclared Economy in FYR Macedonia: A baseline report*, GREY Working Paper no. 3. Sheffield: Sheffield University Management School. 2014.
- Eglash, R. (2011). Multiple objectivity: an anti-relativist approach to situated knowledge, *Kybernetes*, 40(7/8) pp. 995 – 1003.
- Entman, R. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4).

- Eurofound Online (2009a). *Regulations for legalising domestic work, Poland*. Retrieved from <http://www.eurofound.europa.eu>.
- Eurofound Online (2009b). *Information campaign on undeclared wages, Estonia*. Retrieved from <http://www.eurofound.europa.eu>.
- Eurofound Online (2013a). *Unemployed attendance scheme, Czech Republic*. Retrieved from <http://www.eurofound.europa.eu>.
- Eurofound Online (2013b). *Fair Play project, Iceland*. Retrieved from <http://www.eurofound.europa.eu>.
- Eurofound Online (2013c). *Information campaign on tax compliance, Estonia*. Retrieved from <http://www.eurofound.europa.eu>.
- Eurofound Online (2013d). *Illegal Work Leaves Traces campaign, Montenegro*. Retrieved from <http://www.eurofound.europa.eu>.
- European Commission. (2007). *Stepping up the fight against undeclared work*. Brussels: European Commission.
- European Commission. (2001), *European governance — a white paper*. Brussels: European Commission.
- European Commission, 2011, *The former Yugoslav Republic of Macedonia 2011 Progress Report*, COM (2011) 666 final, Brussels.
- European Commission (2013a). *EVALSED: The resource for the evaluation of Socio-Economic Development*. European Commission. Brussels.
- European Commission (2013b). *Communication from The Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Strengthening the Foundations of Smart Regulation – Improving Evaluation*. European Commission. Brussels.
- Ezrahi, Y. (2004). Science and the Political Imagination of Contemporary Democracies. In Jasanoff, S., *States of knowledge: the co-production of science and social order* (p.13-45), London, Routledge.
- Feld L. and Frey B. (2002). Trust breeds trust: How taxpayers are treated. *Economics of Governance*, 3(2): 87–99.
- Franic, J., Williams, C.C. (2014) *Undeclared Work in Croatia: A baseline report*, GREY Working Paper no. 2. Sheffield: Sheffield University Management School.
- Glanville, R. (2009), *Black Boxes, Cybernetics and Human Knowing*, 153-167
- Grim, J and Andsager, J. (2011). Framing immigration: geo-ethnic context in California newspapers. *Journalism and Mass Communication Quarterly*, 88(4).

Grzymaña-Busse A. (2003). Political Competition and the Politicization of the State in East Central Europe. *Comparative Political Studies*, 36(10), p.1123-1147.

Hanberger, A. (2001). What is the Policy Problem? Methodological Challenges in Policy Evaluation, *Evaluation*, 7(1): 45–62.

Hartner M., Poschalko A., Rechberger S. and Kirchler E. (2010). Taxpayers' compliance by procedural and interactional fairness perceptions and social identity. *Journal of Psychology & Economics*, 3(1), p.3-12.

Hedberg, B. (1981). How organizations learn and unlearn. In Handbook of organizational design, ed. P. C. Nystrom and W. H. Starbuck, 3–27. Oxford: Oxford University Press.

Hellmann, T., Perotti, E. (2011). The Circulation of Ideas in Firms and Markets. *Management Science*, 57(10), pp.1813-1826.

Hoerner, J. and Stephenson, P. (2012). Theoretical Perspectives On Approaches To Policy Evaluation In The EU: The Case Of Cohesion Policy. *Public Administration* Vol. 90, No. 3, (699–715)

HM Treasury, (2011a). *The Magenta Book: Guidance for evaluation*. HM Treasury. London.

HM Treasury, (2011b). *The Green Book: Appraisal and Evaluation in Central Government*. HM Treasury. London.

House of Lords (2013). *Behaviour Change*. London : The Stationery Office Limited.

House E., and Howe K. (1999) Values in Evaluation and Social Research. Thousand Oaks, CA: SAGE.

Hutcheon, P. (1970) Sociology and the objectivity problem. *Sociology and Social Research*, 54(2), 153-171.

Jasanoff, S. (2004) Ordering knowledge, ordering society. In Jasanoff, S., *States of knowledge: the co-production of science and social order* (p.13-45), London, Routledge.

Koleva, D. (2015, 29 June). Държавата е похарчила 1.5 млрд. лв., за да повиши заетостта с 2% [The State has spent 1.5 bln leva to raise employment by 2%]. *Sega Online*. Retrieved from <http://www.segabg.com>.

Kraft, M. and Furlong, S. (2010). *Public policy: politics, analysis, and alternatives*, 3<sup>rd</sup> Edition, Washington, D.C. CQ Press.

Lascoumes P., Le Gales P. (2007). Understanding Public Policy through Its Instruments—From the Nature of Instruments to the Sociology of Public Policy Instrumentation. *Governance: An International Journal of Policy, Administration, and Institutions*, 20(1), p.1–21.

Leibniz-Institute fuer Laenderkunde (2015). *За по-добро развитие на градската среда в София: граждански идеи и усилия* [For a better development of the urban environment in Sofia: Citizen ideas and effort], Leibniz-Institute fuer Laenderkunde. Sofia.

- Leicester, G. (1999) Viewpoint: The Seven Enemies of Evidence-Based Policy, *Public Money & Management*, 19(1), 5-7.
- Little, D. (2010) Philosophy of Sociology, in *Philosophies of the Sciences: A Guide* (ed F. Allhoff), Wiley-Blackwell, Oxford, UK.
- Mathison, S. (2005). *Encyclopedia of Evaluation*. Sage Publications: Thousand Oaks.
- Matthews, P. (2012) Problem definition and re-evaluating a policy: the real successes of a regeneration scheme, *Critical Policy Studies*, 6(3), 243-260.
- Matthes, J. (2012). Framing Politics: An Integrative Approach. *American Behavioral Scientist*, 56(3) 247–259.
- McKnight, D., Cummings, L and Chervany, N. (1998). Initial Trust Formation in New Organizational Relationships. *The Academy of Management Review*, 23(3), pp. 473-490
- Meyer-Sahling J. (2004). Civil Service Reform in Post-Communist Europe: The Bumpy Road to Depoliticisation. *West European Politics*, 27(1), p.71–103.
- Meyer-Sahling J., Veen T. (2012). Governing the Post-Communist State: Government Alternation and Senior Civil Service Politicization in Central and Eastern Europe. *East European Politics*, 1.
- Meyer-Sahling J., (2008). The changing colours of the post-communist state: The politicisation of the senior civil service in Hungary, *European Journal of Political Research*, 47, p.1–33.
- McDavid, J., Hawthorn L. (2013). *Program Evaluation and Performance Measurement: An Introduction to Practice*. SAGE Publications, Inc.
- Murphy K. (2003). Procedural justice and tax compliance. *Australian Journal of Social Issues*, 38(3): 379–407.
- Nagel, S. (2001). WIN-WIN methods. In “Handbook of Public Policy Evaluation” ed. Nagel S. Thousand Oaks : SAGE Publications.
- Nonchev, A. (2009). *Monitoring and Evaluation of Public Policies and Programs*. Center for the Study of Democracy. Sofia.
- Nunns, H. (2008). Responding to the demand for quicker evaluation findings. *Social Policy Journal of New Zealand*, 34.
- Onu D. and Oats L. (2014). *Social Norms and Tax Compliance. Discussion Paper: 006-14*. University of Exeter: Tax Administration Research Centre
- Patton M.Q. (1997) *Utilization-Focused Evaluation*. Thousand Oaks, CA: SAGE.
- Pawson, R. (2006). *Evidence-based Policy: A Realist Perspective*. Sage Publications. London.

- Pawson, R., Greenhalgh, T., Harvey, G., Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, 10(1): 21–34.
- Pawson, R., Tilley, N. (2004). Realist Evaluation. In Realist Evaluation. Sage Publications. London.
- Pawson, R. (2002a). Evidence-based Policy: In Search of a Method. *Evaluation*, 8(2): 157–181.
- Pawson, R. (2002b). Evidence-based Policy: The Promise of ‘Realist Synthesis, *Evaluation*, 8(3): 340–358.
- Pearce, W. and Raman, S. (2014). The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance, *Policy Sci* 47:387–402.
- Palenberg, M. (2011). Tools and Methods for Evaluating the Efficiency of Development Interventions. *Evaluation Working Papers*. Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.
- Polak, M., Mishkov, L., and Williams, C.C. (2015). Designing Focus Groups and Experiments to Evaluate Policy Approaches and Measures for Tackling Undeclared Work. GREY Working Paper no. 7, Sheffield University Management School, University of Sheffield, Sheffield
- Rist, R. Introduction (1995) In Policy Evaluation: Linking Theory to Practice, Ray Rist (ed.). Aldershot: Edward Elgar.
- Rocheftort, D. and Cobb, R. Problem definition: an emerging perspective. In In Policy Evaluation: Linking Theory to Practice, Ray Rist (ed.). Aldershot: Edward Elgar.
- Sanderson, I. (2002) Evaluation, Policy Learning and Evidence-based Policy Making. *Public Administration*, 80 (1), 1–22.
- Scott, J. (2000). Rational Choice Theory. In Understanding Contemporary Society: Theories of The Present, edited by G. Browning, A. Halcli, and F. Webster. Sage Publications.
- Slemrod, J., Blumenthal, M., Christiana, C. (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79, 455–483
- Solesbury, W. (2001). *Evidence-based Policy: Whence It Came From and Where It’s Going*. ESRC UK Centre for Evidence Based Policy and Practice – Working Paper 1, London
- Steel, S. 1970. Program Evaluation a broader definition. *Journal of Extension*. Summer 1970.
- The World Bank (2010) *Bulgaria. Ex-post impact assessment of the Act on Limiting Administrative Regulation and Administrative Control on Economic Activity. Better Regulation Series Vol. III*. Washington D.C.: World Bank Group.
- Torgler, B., Schneider, F. (2009). The impact of tax morale and institutional quality on the shadow economy. *Journal of Economic Psychology*, 30, 228–245.

- Torgler B. (2003). Tax morale, rule-governed behaviour and trust. *Constitutional Political Economy*, 14(2): 119–140.
- Torgler B. (2007). *Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis*. Cheltenham: Edward Elgar
- Veselý, A. (2007). Problem delimitation in public policy analysis. *Central European Journal of Public Policy*, 1(1) pp 80–100.
- White, L. (2002). Interactive Policy Analysis: Process Methods for Policy Reform. In Nagel, S. ed. *Handbook of Public Policy Evaluation*, Sage Publications. London.
- Weiss, C. (1979). The Many Meanings of Research Utilization. *Public Administration Review*, 39(5), pp. 426-43.
- Wenzel M. (2004). An analysis of norm processes in tax compliance. *Journal of Economic Psychology*, 25(2): 213–228
- Williams C.C. (2014). Policy approaches towards undeclared work: a conceptual framework. GREY Working Paper no. 4, Sheffield University Management School, University of Sheffield, Sheffield.
- Williams, C.C. (2015). Explaining cross-national variations in the informalisation of employment: some lessons from Central and Eastern Europe. *European Societies*.
- Williams, C.C., Franic, J. (2015). Tackling The propensity towards undeclared work: some policy lessons from Croatia. *South East European Journal of Economics and Business*, 10 (1), 18-31.
- Williams, C.C. and Horodnic, I. (2015a). Evaluating the prevalence of the undeclared economy in Central and Eastern Europe: an institutional asymmetry perspective. *European Journal of Industrial Relations*, doi:10.1177/0143831X14568835.
- Williams, C.C. and Horodnic, I. (2015b). Tackling the Informal economy in South-eastern Europe - an institutional approach. *Southeast European and Black Sea Studies*,
- Williams, C.C. and Martínez, A. (2014). Explaining cross-national variations in tax morality in the European Union: an exploratory analysis. *Studies of Transition States and Societies*, 6(1).
- Williams C.C. and Renooy P. (2008) Measures to tackle undeclared work in the European Union. Dublin: Eurofound.
- Williams, C.C. and Shahid, M. (forthcoming 2015) “Informal entrepreneurship and institutional theory: explaining the varying degrees of (in)formalisation of entrepreneurs in Pakistan”, *Entrepreneurship and Regional Development*, <http://dx.doi.org/10.1080/08985626.2014.963889>
- Williams, C.C., Dzhekova, R., Baric, M., Franic, J. and Mishkov. (2014a). Assessing the cross-national transferability of policy measures for tackling undeclared work. GREY Working Paper no. 5, Sheffield University Management School, University of Sheffield, Sheffield.
- Williams, C.C., Franic, J., Dzhekova, R. (2014b). Explaining the undeclared economy in Bulgaria: an institutional asymmetry perspective. *South East European Journal of Economics and Business* Volume 9 (2), 33-45.

Williams, C.C., Dzhekova, R., Franic, J. and Mishkov, L. (2015a) ‘Evaluating the policy approach towards the undeclared economy in FYR Macedonia’, *International Journal of Entrepreneurship and Small Business*, Vol. 24, No. 2, pp.268–286.

Williams, C.C., Horodnic, I. and Windebank, J. (2015b) “Explaining participation in the informal economy: an institutional incongruence perspective”, *International Sociology*, Vol. 30, No. 3, pp. 294-313.

Williams, M. (2005) Situated Objectivity, *Journal for the Theory of Social Behaviour*, 35(1).