BAM 2017

On the quest for multi-methods in IS evaluation: A Qualitative Comparative Analysis

Completed Research Paper

Abstract

This paper responds to calls for new approaches to IS evaluation. It does this by introducing fuzzy-set Qualitative Comparative Analysis as a new IS evaluation approach that augments the qualitative tradition by supporting cross-case analysis and theory development. Rather than disaggregating cases into independent, analytically-separate variables, fuzzy-set Qualitative Comparative Analysis advocates an approach to IS evaluation which explores the holistic effects of causal conditions working in conjunction with each other. The paper uses qualitative coding procedures and fuzzy-set Qualitative Comparative Analysis in a sequential manner to discover two typologies of monitoring systems success based on automated and manual validations respectively. Theoretical, methodological and practical implications of the use of fuzzy-set Qualitative Comparative Analysis are discussed in the context of a multi-case evaluation of monitoring systems established in the course of the implementation of a major European Union socio-economic support programme.

Key words: IS Success, monitoring systems, qualitative comparative analysis, typologies, evaluation.

Introduction

Despite calls for new approaches to Information Systems (IS) evaluation (Burton-Jones et al. 2015; Kaplan and Shaw, 2004; Petter et al., 2012), few studies have reported rich cross-case comparative research that employs a distinctive configurational framing for cross-case comparisons similar to the analytical procedures used in quantitative research (Ragin, 2008). Doing such cross-case comparative research requires the use of specific techniques to locate and characterise contextual knowledge and from it develop robust, if modest, generalisations that apply to cases beyond the sample under investigation (Berg-Schlosser et al., 2009; Dunleavy et al., 2006). To achieve this goal we introduce in this paper Qualitative Comparative Analysis (QCA) as a new approach to IS evaluation that transcends the quantitative-qualitative methodological divide (Ragin, 2008). Using the increasing body of IS research that advocates set-theoretic methods (El Sawy et al., 2010) and applies them (Henfridsson and Bygstad, 2013; Rivard and Lapointe, 2012; Tan et al., 2016), we show that QCA can serve as a comparative case study evaluation methodology that employs counterfactual analysis and extends beyond the replication logic which informs most extant comparative approaches (Cavaye, 1996; Eisenhardt, 1989; Yin 2014). While replication attempts to either confirm or disconfirm findings from across multiple cases, QCA refines this quasi-experimental logic using counterfactuals to simplify “complexity in a theoretically-guided manner” (Ragin, 1987: 83). QCA is particularly suitable for IS evaluation research because it extends the static, variance-oriented logic to include conjunctural and heterogeneous causation. It addresses those situations where “different conditions combine in different and sometimes contradictory ways to produce the same outcome” (Ragin, 2000: 40). QCA can generate new insights into causal arguments that can serve as a basis for theory development from multiple case studies (Berg-Schlosser et al., 2009). Theory development is valuable in IS evaluation when it shifts attention from the assessment of the “net effects” of causal variables to a more contextual understanding of the multiple possible ways causal conditions may combine to produce a given effect.

In this paper, we interweave QCA with process-tracing methods to chart the temporal pathways leading to the outcome of interest (i.e., positive impact) and develop typological theories that capture different types of positive impact stemming from different configurations (or recipes) of causal conditions. Rather than studying the “net effects” of causal conditions working independently of each other, our approach examines the multiple ways that conditions may combine to produce the outcome of interest. The empirical study reported here applies QCA in this spirit using a sequential, multi-method research strategy (George and Bennett, 2005). The cases we study are national monitoring systems established and run by governments in the course of their implementation of a major European Union socio-economic support programme. These systems are complex socio-technical configurations that present technical, legislative, organisational and social features. Our research question is then: *What aspects of these monitoring systems are relevant for a positive effect to be seen (be it described as an outcome, consequence or impact) and in what configurations may such aspects combine to produce these effects?* In brief, our approach uses case study data and qualitative coding procedures informed by prior theory to inductively derive plausible causal conditions and the outcome of interest (Ragin, 2000). We then use QCA and its fuzzy-set variant to identify and refine complex cross-case patterns. By sequentially combining qualitative coding procedures with set-theoretic methods, we implement a research strategy that uses in-depth case knowledge but is also able to offer distinctive cross-case insights within the wider European monitoring context. This approach serves to unravel the multiple, dynamic and asymmetric ways by which configurations of causal conditions combine to produce the presence (or absence) of the outcome of interest. The remainder of this paper unfolds as follows. Section two reviews distinct approaches to IS evaluation. Section three explains our research setting and strategy by outlining the details of our data analysis technique. Section four and five sketch our coding and calibration procedures with regard to our causal conditions, i.e., information and system quality, and outcome, i.e., impact, respectively. Section six presents a configurational analysis of the causes under investigation to pinpoint the causal recipes for positive and non-positive impacts. Section seven discusses our findings and concludes this work.

Theoretical approaches to IS evaluation

Evaluation has long been understood as essential for advancing the IS field, interweaving socio-technical, cultural, economic and business issues in an effort to understand, measure and assess the worth, value or success (DeLone and McLean, 1992; 2003) of IS activities. Evaluation of IS is often conceived as a judgment of worth and draws on diverse theoretical perspectives (Burton-Jones et al., 2015; Kaplan, 2001; Klecun and Cornford, 2005; Klecun et al., 2014). Early IS evaluation research was predominantly based on variance models, employing laboratory and field experiments, as well as quasi-experimental survey designs (Van de Ven, 2007). Whether by manipulation through experimental design or by repeated observation of naturally-occurring events typically measured using surveys, variance models are informed by the assumption that the precursor (X) is a necessary and sufficient condition for the outcome (Y) (Markus and Robey, 1988). This form of research design strives for universality and hence generalisability. Experimentation tries to minimise the differences (except one - the intervention) between experimental and control groups, stripping context away and yielding refined and universal results. To critics, however, these results are valid only in other "context-less" situations (Pawson and Tilley, 1997). Likewise surveys analyse presumed causal associations between independent and dependent variables in a (random, possibly stratified) sample of observations drawn from a coherent sampling frame (Van de Ven, 2007). By controlling for plausible spurious or confounding variables, surveys attempt to draw inferences from the sample to the target population, assuming that the random sample of observations and the population are "like objects", that is comparable, substitutable, independent instances of "the same thing" (Ragin, 2000: 48). Over time the assumptions of both the experimental and quasi-experimental approaches to evaluation have been challenged. For example, there have been calls for contingent approaches to evaluation since what "works" in one context may not work in a different context (Pawson and Tilley, 1997).

Spurred by these two perspectives on IS evaluation, that is, Experimentalist vs. Contextualist approaches, IS scholars have chosen different paths. On the one side, are broadly quantitative scholars who assess the value or success of ICT investments with such variables (or metrics) as inter-organisational performance (e.g., Rai et al., 2006), organisational performance (e.g., Devaraj and Kohli, 2003), systems acceptance (e.g., Davis et al., 1989), acceptance and satisfaction (e.g., Wixom and Todd, 2005) and use (e.g., Goodhue et al., 2000), to name but a few. On the other side, are qualitative scholars who conduct in-depth studies exploring the use of IT evaluation methods (e.g., Nijland and Willcocks, 2008), content, context and process issues (e.g., Symons, 1991), as well as pre-and post-implementation issues (e.g., Serafeimidis and Smithson, 2003) within (e.g., Dobson et al., 2007) and across settings (e.g., Uwizeyemungu and

Raymond, 2009). Informed by what many have traditionally seen as incommensurable research paradigms, i.e., quantitative vs. qualitative, and alternative impact evaluation models, i.e., variance vs. process models (Pare´ et al., 2008), these two strands of research have evolved in separate directions (Chan, 2000) in spite of early attempts to integrate them (Kaplan and Duchon, 1988; Gable, 1994; Mingers, 2001) and hybrid models attempting to interweave variance and process theories (Burton-Jones et al., 2015; DeLone and McLean, 1992; 2003). More recently, a new wave of pluralist IS evaluation approaches has highlighted the concerns of multiple stakeholders, the need for cross-case comparative studies and the incorporation of 'soft issues' (Kaplan and Shaw, 2004). In this paper, we seek to join this new pluralist wave of evaluation scholars by introducing QCA as an approach to IS evaluation that transcends the quantitative-qualitative boundaries (Ragin, 2008). QCA extends the logic of variance-oriented models by analysing how social and technical components react to one another in a conjunctural fashion so much so that their properties in combination are different from their properties in isolation (Davis et al., 1992). Thus, by analysing the holistic effects of causes working in conjunction with each other, QCA transposes the context-sensitive logic of qualitative research into the analytic framework of variable-oriented research (Ragin, 1987). In what follows, we corroborate this contention using the original DeLone and McLean's work as our focal model in the dialogue between theoretical ideas and empirical evidence. We choose the DeLone and McLean's (1992) model for two reasons: first, IS scholars have tested this model extensively from a variance-oriented perspective, thus neglecting its configurational (or systemic) features; second, the logic underpinning this model fits our empirical setting quite well and demonstrates how raw monitoring data are transformed into validated information along the monitoring process, thus producing efficient or inefficient results accordingly. We are also heartened by various admonitions from DeLone and McLean themselves to recognise "success as a process construct which must include both temporal and causal influences in determining IS success" (Ibid: 83; emphasis in original; see also DeLone and McLean 2003: 15-16).

Research setting and strategy

The work reported in this paper is motivated by an evaluation to assess the relative success of various monitoring systems established by national government bodies for European Social Fund (ESF) activities in the 2000-2006 programming period. The ESF is the principal European Union (EU) policy instrument to achieve economic cohesion. It operates through a set of programmes in all EU countries funded by the European Commission. It had a budget of approximately 60 billion Euros over the period 2000-2006. The ESF is focused on a consistent set of priorities or objectives (e.g., matching labour market demand and supply, creating net jobs, etc.) and funds corresponding measures or projects (e.g., training projects). As part of the implementation of this policy, the various ESF-funded interventions in each EU country are subjected to monitoring to ascertain the veracity of the claims for funding with regard to what is being achieved and how much money is being invested. Our research is concerned with evaluating the relative success of these national monitoring systems across 7 European country cases (i.e., Austria, England, Flanders, France, Germany, Greece and Hungary) which were purposefully selected to compare their monitoring systems.

These monitoring systems support the inspection and audit of ESF-funded projects in specific ways including their inputs, outputs, results and impacts. For example, for a project providing training (the most common type), data is required on the amount of financial resources used (i.e., inputs), the number of training hours delivered (i.e., outputs), the number of successful trainees (i.e., results) and the long-term indirect effects in the economy or labour market (i.e., long-term impacts). Based on a common schema established centrally by the European Commission, ESF-funded projects are monitored by means of physical and financial indicators, the former referring to synthetic summary metrics tracking project and programme implementation (e.g., number of training places, number of training hours, number of successful trainees, etc.), the latter referring to financial resources used to implement projects (e.g., financial inputs, eligible costs, etc.). More specifically, training providers (or project managers) are expected to enter physical and financial data in their IT systems. These data are then transmitted to Beneficiary Organisations for the purpose of validation (and cross-checking) of data entries. Next, Beneficiary Organisations relay such data to Managing (or Paying) Authorities and their Monitoring Committees who aggregate them and transmit them to the European Commission.

Research strategy and data analysis technique

Due to the lack of theorising about the monitoring and reporting procedures enacted by various European Member States to obtain ESF funding, we used a two-pronged strategy interweaving qualitative coding procedures with QCA in a sequential fashion (George and Bennett, 2005). In the first stage, country experts gathered primary data through semi-structured interviews informed by pre-existing themes. During this stage, data collection and data analysis took place simultaneously leading to the development of an in-depth case study for each country with emerging codes that explained why a monitoring system was deemed as successful or not. This within-case analysis was subsequently corroborated with backward coding to re-examine previous events for overlooked relevance and distil patterned similarities and differences across cases. Again new codes were generated which required the gathering of new data (mostly secondary) to validate them. Through an iterative dialogue between our fledging theoretical ideas and the empirical data, we derived a large number of plausible causal and outcome variables in an inductive fashion as discussed below. For each causal condition and the outcome variable we obtained a "theoretical ideal" (e.g., a country that captures all types of indicators, a country that uses fully-automated monitoring systems, etc.). We used these "ideals" as a yardstick for our calibrations (i.e., the assignment of fuzzy-set membership scores for each country).

In the second stage, these inductively-derived causal and outcome variables were subjected to a fuzzy-set Qualitative Comparative Analysis (fsQCA) to reveal complex cross-case understandings of the type of impact and, indirectly, success. Compared with its crisp-set variant (which uses binary variables), the fuzzy-set approach allows for degrees of set membership to be specified and thus captures more nuanced causal relationships. For example, rather than dichotomising the impact variable as positive or negative, it allows for different degrees of membership in the impact variable, thus capturing more fine-grained distinctions. By sequentially moving from within-case analysis to a fsQCA strategy of cross-case analysis, we endeavoured to identify commonalities within the same types of cases and differences across distinct types of cases (Ragin, 1987). During this second stage, our analysis proceeded in four steps (Cf. Schneider and Wagemann, 2012: 178-193).

Step 1: calibrating data and converting the distribution matrix into the truth table.

After causal and outcome variables were calibrated (see Tables 1, 2, & 3), the calibration scores were used to construct a truth table that lists all logically-possible combinations of causal conditions (see Table 5). Next, each country case was assigned to that combination of conditions in which its membership exceeded 0.50, thus obtaining non-remainder (i.e., populated) and remainder (i.e., empty) rows.

Step 2: determining the outcome value for each truth-table row

In the second step, we derived the outcome value for each truth-table row. This was based first on setting a minimum number-of-cases threshold for each row. Having a small number of cases, we set the minimum number of required cases in a row at 1. We then considered the consistency level for each row populated with cases. Consistency (sufficiency) is a measure that gauges the degree to which a causal condition or combination of conditions is a subset of the outcome. We set the lowest level of acceptable consistency at 0.90, above the minimum recommended threshold of 0.85 (Cf. Ragin, 2008: 136), Accordingly, truth-table rows with consistency above 0.90 were given a score of 1 since the configuration was a consistent subset of the outcome, or 0 if not. The empty rows in the truth table were labelled as "remainders" because they did not meet our minimum number-of-cases threshold.

Step 3: counterfactual analysis

The third step uses counterfactual analysis based on Boolean logic. Through pencil-and-paper procedures, we produced new combinations of causal conditions drawing on 'what if' claims about the outcome of cases that did not occur in the sample (i.e., the 'remainders'). In this way we developed a theoretically-enhanced solution. Essentially, we asked whether the outcome of interest (i.e., positive impact) would be present if there were cases that populated the empty rows of the truth table. Using both our substantive knowledge of individual cases and the theoretical knowledge derived from the DeLone and McLean model, we obtained two solutions (or recipes for positive outcomes) from our data:

1) a conservative solution where no 'remainders' were included and which only drew on empirical data. However, based on substantive and theoretical knowledge we moved beyond the constraints of this solution (see Analysis Section below);

2) a theoretically-enhanced solution that included all counterfactuals from the pool of 'remainders' on the assumption that they exhibited the outcome of interest. These counterfactuals permitted the inclusion of both simplifying (IQ*SQ) and non-simplifying (IQ*sq) remainders to help us move beyond the empirical domain on the basis of theoretical plausibility (see Tables 6 & 7).

Step 4: Interpreting findings

Last, we returned to the cases to trace the causal processes linking explanatory conditions with the outcome of interest. In this final step, we used the theoretically-enhanced solution terms as scope conditions for unravelling more granular causal recipes and for identifying both core (necessary) and peripheral (contingent) conditions indicating different types of successful monitoring systems (see Table 8). All findings were finally validated with the fsQCA 2.5 programme.

Information and system quality calibrations

Our qualitative coding procedures were informed by prior theory, i.e., the DeLone and McLean model of IS success. Based on this theory (or model), we identified broadly-interrelated themes by assuming that information and system quality contributed to positive impact either jointly or separately. Based on these pre-existing themes, we developed more nuanced conceptualisations through a back-and-forth cycling between theoretical ideas and empirical evidence (Ragin, 2000), thus progressively refining theoretical and substantive knowledge and generating more fine-grained components of information and system quality. The information quality (IQ) variable captured the various facets of the indicators, the system quality (SQ) variable represented the technical features of the monitoring systems and the impact variable the consequences (or effects) stemming from the use of the monitoring system conceived as a socio-technical system. Subsequently, for each information or system quality component and for our outcome, i.e., impact, we inductively derived a "theoretical ideal" as "the best imaginable case in the context of the study that is logically and socially possible" (Basurto and Speer, 2012: 166). We then used these "ideal types" as external standards against which our empirical cases could be calibrated (Basurto and Speers, 2012; Goertz and Mahoney, 2012). Based on Ragin's (2000: 168) advice to use five-value schemes when data are too weak to support fine-grained distinctions, we ranked the country cases at hand in the following fashion:

• fully shares the inductively-derived criteria of the ideal type, scored as 1;

• meets the inductively-derived criteria of the ideal type in a substantive fashion, scored as 0.75;

• partially meets some of the inductively-derived criteria of the ideal type while lacking other crucial criteria, scored as 0.5 (i.e., the crossover point);

• only meets a minority of the inductively-derived criteria of the ideal type, scored as 0.25;

• does not meet the inductively-derived criteria of the ideal type, scored as 0.

Information quality components

To arrive at the IQ components we progressively gathered more focussed data and formulated less vague conceptualisations. By iteratively re-assembling data that were fractured during open coding (Strauss and Corbin, 1998) we arrived at more abstract IQ dimensions, namely comprehensiveness, consistency and currency. Comprehensiveness refers to the scope of the information being collected which needs to be wide enough to meet the needs of various stakeholders. Consistency refers to the use of common definitions in compliance with the European Commission requirements. Currency indicates the regular updating of monitoring data and indicators in accordance with changing information needs. Primary and secondary data were coded into categories and sub-categories. As coding progressed, we sought the "full range of variation among the phenomena under scrutiny" (Corbin and Strauss, 1990: 13) to unearth deeper relationships between the data and subsequently formulate short descriptions that summarise the information stemming from several sources in one statement that best reflects the case (Basurto and

Speer, 2012). Based on such summary statements, we transformed qualitative data into fuzzy-membership scores by matching each country case statement against its “ideal types” (Goertz and Mahoney, 2012). Table 1 below shows the summary statements associated with each country, as well as its fuzzy-membership scores along the three dimensions of IQ.

Table 1. Fuzzy-set Membership Scores for Information Quality Components

Country Case	Summary Statement	Comprehensiveness Score	Consistency Score	Currency Score
Austria	Financial and physical indicators used but no availability of data covering the employment needs of prospective job seekers. Consistent hierarchy of input, output and result indicators which are regularly updated to meet changing needs.	0.75	1.00	1.00
England	Financial input indicators but no physical output indicators and ambiguous project results. Inconsistent indicators due to conflicting definitions across Regional Government Offices, as well as contradictory requirements with the European Commission. No systematic updating of indicators during programming period despite attempts at doing so.	0.50	0.25	0.25
Flanders	Input, output and result indicators but no strategic indicators. Misalignment between ESF and Flemish definitions due to divergent policy cycles. Fragmentation between physical indicators. Data updated in a regular fashion but out-of-date indicators concerning target groups.	0.75	0.25	0.75
France	Financial input and physical output indicators absent in the first three years of the programming period under investigation but guidelines enacted thereafter. No recording of result and impact indicators because “exhaustiveness is unrealistic”. Consistent definitions have been developed more recently by issuing a guide for users but its application is still at an embryonic stage. Mixed picture with regard to updates as some indicators are updated regularly while others are neither recorded nor updated.	0.50	0.50	0.50
Germany	Input, output and result indicators but no long-term impacts. Serious	0.75	0.25	0.75

	fragmentation of physical result indicators across States. Data updated regularly but new action plan for indicators has slowed down the introduction of new indicators.			
Greece	Input, output and result indicators but no global objectives. Fragmentation of physical output and result indicators. Regular updating of indicators in accordance with new information needs but their fragmentation has slowed down this process	0.75	0.25	0.75
Hungary	Input indicators but lack of impact indicators and poor availability of physical output and result indicators. Lack of coherent definitions across programmes, projects and objectives in terms of output and result indicators. Out-of-date information about the scope of indicators and their target values.	0.50	0.25	0.25

System quality components

Although system quality components reflect the more engineering-oriented performance characteristics of the monitoring systems (Cf. DeLone and McLean, 1992: 64), we broadened our analysis to their associated social processes (e.g., verification and validation procedures, consistency checks, etc.) because technical systems are enacted by following specific monitoring routines that are embedded within structures and broader socio-institutional contexts (Cornford et al., 1994). Once again, by using an iterative dialogue between theoretical and practical knowledge, we unravelled several codes from the empirical data. By bringing related codes together, and using prior theory, we identified more abstract system quality dimensions, namely compatibility, reliability and automation. Compatibility refers to systems relying on common communication protocols that can effectively and efficiently transmit data. Compatibility is the technical equivalent of consistency, the former referring to technical standards and the latter to business-level definitions concerning the system of indicators. Reliability instead refers to the dependability of a system, component and/or monitoring procedure. Lastly, automation refers to the degree to which the procedures enacted for the verification and validation of data are performed by a pre-programmed system. Using the same calibration procedures outlined above, we obtained the fuzzy-membership scores for the system quality dimensions for each case. Table 2 below outlines these calibrations.

Table 2. Fuzzy-set Membership Scores for System Quality Components

Country Case	**Summary Statement**	**Compatibility Score**	**Reliability Score**	**Automation Score**
Austria	Fully-compatible networked system environment across regional, state and federal levels where data entries are based on a specific table blueprint (and XML schema) but lack of structured data underpins exchanges with training providers. Reliable and resilient systems in	0.75	0.75	0.50

	spite of occasional data losses. Manual validation routines (i.e., "four-eye principle") within Beneficiary Organisations but ongoing implementation of SAP module for automated monitoring purposes.			
England	Information transmissions based on interactive disks. No TCP/IP standards. No tracking of audit trails and control procedures due to systematic failures. Heavy use of manual procedures for data validation purposes across Beneficiary Organisations.	0.25	0.25	0.25
Flanders	No transfer of structured data between Client Following System (CFS) and Microsoft Suite but TCP/IP standards in place. The CFS is highly dependable but the Microsoft Suite suffers from minor glitches. Pronounced use of manual procedures for data validation purposes despite data convergence documents enabling full coherence between and among indicators.	0.50	0.75	0.50
France	No structured transmissions of electronic data due conflicting data standards but systems' interfaces have been laid out. The Managing Authority's system (Application FSE) suffers from frequent breakdowns that are exacerbated by functionality problems and softer issues concerning the certification of expenses. Manual verifications due to technical limitations of Application FSE with regard to the aggregation of regional data.	0.50	0.25	0.25
Germany	Networked environment based on structured input/output data (i.e., templates), XML standards and client-server architectures but the ongoing renewal of servers has undermined systems connectivity to a degree. Fully-dependable monitoring system. Built-in plausibility checks, colour-code system and automated generation of templates but manual cross-checks aimed at verifying data entries and ascertaining that financial claims match actual costs.	0.75	1.00	0.75

Greece	Interoperability issues upstream (Managing Authority) with regard to physical data but seamless retrieval of financial data from the Data Warehousing System. Heavy use of electronic-mail transmissions downstream (Training Providers). Legacy issues between new and old electronic archives stored in the Data Warehousing System. This system fails repeatedly. It does not perform summations of physical data and decimal inputting, thus triggering manual aggregations. Physical data are manually validated. However, data mining tools automatically flag up financial errors.	0.50	0.25	0.50
Hungary	Shared standards based on TCP/IP protocols but ongoing development of new modules in the Unified System has hindered the seamless transmission of information across interfaces. Unclear picture with regard to system failures as despite official reports maintaining that the Unified System is reliable, interviewees have stated that the system is slow, subject to breakdowns and unable to save tables properly. Numerical data types automatically checked but the same does not hold for textual data.	0.50	0.50	0.50

Calibrating impact

Using our existing knowledge on impacts, consequences or outcomes (e.g., Markus and Robey, 2004; Seddon, 1997) as a sensitising device, we extracted themes from our primary and secondary data that indicated positive impact or the lack thereof (e.g., error-free and timely data delivery, minimal or no duplication of work, satisfied stakeholders, etc.). By combining these themes, we conceived of our outcome of interest, i.e., positive impact, as the intended and/or unintended effects on organisational constituencies stemming from the enactment of the monitoring system. Though we used multiple sources to triangulate our data, we adopted the perspective of Managing/Paying Authorities, Monitoring Committees and Beneficiary Organisations to assess these effects because of the mandatory nature of the monitoring system. Thus, we formulated judgments from the perspective of relevant stakeholders about what is valuable from capturing and sharing validated monitoring data efficiently, and, conversely, what is not valuable from collecting, transmitting and validating monitoring data inefficiently. Subsequently, we derived 'ideal types' for positive cases and formulated summary statements that captured the evidence concerning every single case in the best possible way. By repeating the same coding procedure outlined above, we obtained the calibrations shown in Table 3.

Table 3. Fuzzy-set Membership Scores for Positive Impact

Country Case	Summary Statement	Fuzzy-set score for

		positive impact
Austria	Mostly positive effect on participants thanks to a concerted monitoring approach revolving around compatible and reliable IT systems. However, the 'four-eye' principle has triggered occasional delays in terms of data delivery, some data processing errors and duplication of manual controls (i.e., unnecessary costs).	0.75
England	Little or no positive effect on participants because of conflicting perceptions of information requirements (e.g., widespread lack of satisfaction across stakeholder organisations worsened), manual data re-keying (e.g., numerous re-keying errors, duplication of costs, etc.) and manual validations of inconsistent data (e.g., undetected errors, data delivery delays, etc.).	0.25
Flanders	Instances of both non-positive effects (e.g., duplication of work stemming from submitting reports in accordance with both ESF and Flemish definitions) and positive effects (e.g., data convergence documents facilitating data aggregation & disaggregation). Quality control reports may trigger undetected errors. Yet they nurture good bilateral 'coaching' relations between Project Managers and Partner Institutions, thus generating deep knowledge of ESF projects.	0.50
France	Little or no positive effect on participants stemming from monitoring requirements perceived to be too bureaucratic and not geared towards a common purpose. Manual validations of inconsistent data entail aggregation & disaggregation errors, as well as undetected errors. There are also data losses stemming from frequent breakdowns, undetected errors caused by softer certification issues, as well as unnecessary delays in the delivery of monitoring data and re-keying errors.	0.25
Germany	Mostly positive effect on participants thanks to a standardised approach for the collection, storage and transmission of monitoring data leading to more efficient coordination between Federal and State monitoring activities and to seamless verification and transmission of monitoring data. However, overwhelming data requirements and inconsistent result indicators have led to minor errors in the aggregation & disaggregation of monitoring data at the Federal level.	0.75
Greece	Little or no positive effect on participants stemming from technical limitations of the Integrated Data Warehousing System causing additional workload and systematic failures. Manual validations of inconsistent physical data trigger aggregation & disaggregation errors, undetected errors and data processing delays. There are also additional monitoring costs because of Beneficiary Organisations' opportunistic behaviours.	0.25
Hungary	Little or no positive effect on participants stemming from inconsistent indicators, lack of interoperability across interfaces, tensions between standardisation and fragmentation, slow IT systems unable to save tables properly and manual validations of textual data.	0.25

Analysis

This Section presents the fsQCA analysis of the cases described above by conceiving of them as configurations (or combinations) of explanatory conditions (Fiss, 2011). In coding and calibrating the data, we combined sets of causal dimensions that go together conceptually into smaller groups of conditions. These were then re-conceptualised at a higher level of abstraction in terms of IQ and SQ. We now assess the degree of membership in each higher-order construct using the rule of the minimum which goes hand-in-hand with our in-depth case knowledge. This rule assumes that a case with weak membership in one of its constructs' components could have, at best, only weak membership in any combinations of dimensions that include this specific component (Ragin, 2008). Table 4 shows the resulting fuzzy-membership scores for IQ and SQ.

Table 4. Aggregated Fuzzy-set Membership Scores for Information & System Quality

Country Case	**Fuzzy-membership scores for Information Quality (IQ)**	**Fuzzy-set score for System Quality (SQ)**
Austria	0.75	0.50
England	0.25	0.25
Flanders	0.25	0.50
France	0.50	0.25
Germany	0.25	0.75
Greece	0.25	0.25
Hungary	0.25	0.50

We then create a truth table that lists all logically-possible combinations of causal conditions whether they are populated with cases or not. Setting the minimum number-of-cases threshold at 1 and establishing the lowest level of acceptable consistency at 0.90, Table 5 shows as fuzzy subsets of the outcome those combinations of causal conditions whose consistency score is at or above the cut-off value of 0.90.

Table 5. Truth Table for Positive Impact

IQ	**SQ**	**Number of cases with score more than 0.50**	**Outcome code (based on consistency score)**	**Consistency (Xi≤ Yi)=∑[min(Xi, Yi)]/∑(Xi)**	**PRE(Xi≤ Yi)=∑[min(Xi, Yi)]- ∑[min(Xi, Yi, yi)]/∑(Xi)- ∑[min(Xi, Yi, yi)]**	**Product**
0	0	2	0	2/3.5= 0.57	0/1.5=0.00	0.00
0	1	1	1	2.5/2.75= 0.91	0.50/0.75= 0.67	0.61
1	0	0	Remainder	2/2.25= 0.89	0.25/0.50= 0.50	0.44
1	1	0	Remainder	2/2= 1	0.25/0.25= 1	1.00

Notes: Min indicates the selection of the lower of the two values; Xi designates the degree of membership in the combination of causal conditions; Yi designates the degree of membership in the outcome; ≤ designates sufficiency (i.e., "is sufficient for") or subsethood (i.e., "less than or equal to"); > designates "more than"; ∑ designates summation; Consistency= degree to which membership in that corner of the vector space is a subset of membership in the outcome; PRE= proportional reduction in error calculation

Product= the multiplicative product of consistency and PRE (note that high product values entail sufficiency"); Remainders designate configurations with no empirical cases with membership scores above 0.50; 0= absence; 1= presence (only applied to IQ and SQ values).

By inspection, Table 5 shows that the combination of non-Information Quality (iq) AND System Quality (SQ) is sufficient for positive impact (see row 2). At this juncture there are two possibilities to pursue. The first and more conservative option is to avoid 'remainders' altogether and stick with the aforementioned solution. However, in-depth knowledge of the empirical data shows that Germany is far from being a good instance of Non-Information Quality (iq) and System Quality (SQ) because, although it is closer to the iq*SQ corner of the vector space, its standardised approach for the collection, storage and transmission of monitoring data compensates to a degree for the fragmentation of result indicators. In other words, it is much easier to aggregate fragmented physical result data if they are stored and transmitted to the Federal Managing Authority in a standardised fashion through the 'template system'. Thus, based on our in-depth case knowledge, we can move beyond the constraints of this conservative solution. The alternative is to incorporate all 'remainders', that is, both IQ*SQ and IQ*sq, in order to yield a theoretically-enhanced solution (Cf. Schneider and Wagemann, 2012: 211-212) as follows (see Table 6).

Table 6. Incorporation of all Remainders
iq*SQ + IQ*SQ + IQ*sq ⟶ Impact Where: iq= Non-Information Quality; IQ= Information Quality; sq= Non-System Quality ; SQ= System Quality

Notes: Logical AND designated by the asterisk (*); Logical OR designated by the plus sign (+); Sufficiency designated by the arrow running from the sufficient conditions towards the outcome (⟶).

By using Boolean algebra (Cf. Ragin, 2008: 156), it follows (see Table 7 below).

Table 7. Theoretically-Enhanced Solution
SQ (iq + IQ) + IQ (SQ + sq) ⟶ Impact SQ + IQ ⟶ Impact

Thus, our theoretically-enhanced solution suggests that positive impact emerges either when SQ is present (e.g., Germany) or when IQ is present (e.g., Austria). To uncover more granular and temporally-meaningful causal recipes, we develop our solutions within these scope conditions. Returning to the cases, we interpret Germany to be a strong instance of Comprehensiveness AND Currency AND Compatibility AND Reliability AND Automation. We also interpret Austria to be a strong instance of Comprehensiveness AND Consistency AND Currency AND Compatibility AND Reliability, thus producing two distinct configurations as follows (Cf. Berg-Schlosser et al., 2009: 16).

Table 8. Causal Configurations for Achieving Positive Impact

Temporal dynamics	Solution$_1$ (Automated validation/Germany)	Events	Solution$_2$ (Manual validation/Austria)	Events
T$_1$	Comprehensiveness* Currency* Compatibility* Reliability* Automation	(Data Entry, Validation, Transmission)	Comprehensiveness* Currency* Compatibility* Reliability* Consistency	(Data Entry, Transmission)
T$_2$				(Manual Validation)
T$_3$	Positive Impact (Timely delivery of accurate data and ESF funding disbursement)			
T$_4$			Positive Impact (Delivery of accurate data and ESF funding disbursement)	
Consistency (sufficiency)	1.00		1.00	
Raw coverage	0.92		0.75	
Unique coverage	0.25		0.08	
Overall solution consistency	1.00			
Overall solution coverage	1.00 (out of which 0.67 is overlapping coverage)			

Notes: Logical AND designated by the asterisk (*); Sufficiency designated by the arrow (→): Temporal sequence designated by the dashed arrow (⇢).

Next, we identify the causally-relevant commonalities between Germany and Austria as the causal core and we label the remaining causes as being part of the periphery. This reveals two recipes for positive impact. To ensure an efficient monitoring process, it is necessary that a monitoring system collects a comprehensive and up-to-date range of financial and physical indicators and relies on dependable technologies with well laid out communication protocols. However, this is not enough. Raw monitoring data must be validated to verify that they have been entered correctly in the system whether at the individual or aggregate level. Two procedures seem possible for this purpose: either validation in an automated fashion or manual validation procedures. While the former procedure may guarantee a speedy and seamless processing of monitoring data at the early stages of the monitoring process, the latter puts a strain on data processing speed and accuracy because human beings are slower and less accurate than technologies in such information processing tasks. Hence, Member States using manual validation procedures must deploy a consistent set of indicators if they are to keep up with the speed and accuracy that can be achieved by technological automation. The Austrian case is telling in this respect. Data transmitted to Beneficiary Organisations are manually cross-checked by at least two colleagues per office through the 'four-eye' principle and, once entered in the Data Warehousing System, they are sent to the Federal Managing Authority where they undergo further checks. The overall effectiveness of manual validation procedures, therefore, depends on consistent definitions of monitoring data and indicators. The more consistent the set of indicators, the faster and more accurate the manual checking and cross-checking of data entries can be.

Conversely, the lack of positive impact (i.e., the failure of the monitoring system) presupposes no compatible communication protocols coupled with inconsistent indicators and non-automated validations. In other words, the manual verification of inconsistent data in conjunction with the re-keying of data entries is a guaranteed recipe for failure of the monitoring system. These asymmetric findings are outlined below.

Table 9. Causal Configurations for Achieving Negative Impact.

Temporal dynamics	Solution$_3$ (Lack of socio-technical standards AND lack of automation; e.g., Flanders)	Events
T_1	Non-Consistency* Non-Compatibility* Non-Automation	(Data Entry, Transmission, Validation)
T_2		(Reiterated data Entry, Transmission, Validation)
T_n	Non Positive Impact (No ESF funding)	(No delivery of accurate/timely data)
Consistency (sufficiency)	1.00	
Raw coverage	0.75	
Unique coverage	0.75	
Overall solution consistency	1.00	
Overall solution coverage	0.75	

Notes: Logical AND designated by the asterisk (*); Sufficiency designated by the arrow (→); Temporal sequence designated by the dashed arrow (⇢).

Discussion and conclusion

Spurred by calls for new directions in evaluation research (Burton-Jones et al., 2015; Kaplan and Shaw, 2004; Petter et al., 2012), we have introduced and applied here a new evaluation methodology, one that we argue opens up new possibilities for evaluation studies and will help evaluation researchers move beyond the constraints of experimental designs by tracing the processes that link configurations of causal conditions with the outcome of interest. Drawing on QCA, and in particular the fuzzy-set variant, this approach looks beyond narrow choices of positivist vs. interpretive, quantitative vs. qualitative, variance vs. process models. As Ragin (1987: 84) proposes, QCA "integrates the best features of the context-oriented approach with the best features of the variable-oriented approach". We would add that QCA should not be applied as the only data analysis technique but should be integrated with other methods (Schneider and Wagemann, 2012). In this paper, we have combined QCA with process-tracing methods (George and Bennett, 2005) to chart temporal pathways that lead to positive impact. Accordingly, we have discovered two distinct types of successful monitoring systems with their associated temporal pathways. We believe that this paper does offer a new meta-model of evaluation to reinvigorate debate on the 'quest for the dependent variable' and in particular its expression in causal processes. Thus, we see this paper as contributing to theory, methodology and practice.

Theoretically, the use of QCA in combination with process-tracing methods has allowed the D&M model of IS success to be re-considered, as its authors intended, as a model with a strong process dimension. Though the D&M model has been extensively tested from a statistical (variance) perspective (e.g., Sedera et al., 2013), few, if any, such tests have identified or addressed the processes and causal links through which organisational impact emerges. In this paper, we demonstrate that, by deploying fsQCA and by tracing the temporal links embedded within individual cases, we can start to develop typological theories – in our case of monitoring systems success – which apply to other similar cases that share a reasonable number of characteristics with the evaluated cases (Berg-Schlosser et al., 2009). Typologies consist of configurations of core and peripheral parts, with the core elements being essential (or necessary) and the peripheral elements being less important and perhaps even "expendable or exchangeable" (Fiss, 2011: 394). This means that IS success can be reconceived as a concept that is neither additive nor multiplicative (Polites et al., 2012) but rather substitutable, where one type of successful monitoring system can substitute for another in a revised temporality. While manual validations may substitute for automated validations, it turns out that the latter are a more powerful typology because they foreshadow the move towards monitoring systems acting as "early-warning systems" (Tödtling-Schönhofer et al., 2011: 13/26).

Alongside and interwoven with the theoretical contribution, this work has methodological and practical implications. Methodologically, the study of impact and, indirectly, IS success is a challenging endeavour because it entails the use of some form of experimental design. We submit that QCA in general and fsQCA in particular provide IS evaluation researchers with an approach that extends beyond any simple replication logic. Where replication looks for patterns of invariance from across multiple cases, QCA searches for multiple, partially-overlapping recipes for the outcome of interest that consist of both systematic (core) and non-systematic (contingent) patterns. Furthermore, complementing QCA with process-tracing techniques is a useful approach because it integrates static cross-case comparisons with more dynamic, process-centred within case analysis. The implications for practice we draw relate to issues of asymmetry. In this paper, we show that the configurations leading to monitoring systems success and failure are asymmetric. While successful monitoring systems presuppose dependable technologies with well-defined communication protocols able to capture and transmit a comprehensive and up-to-date range of indicators, unsuccessful monitoring systems require the lack of socio-technical standards and manual validations. Thus, the removal of a required condition for failure (e.g., removing compatibility constraints to ensure good communication standards) may be a necessary condition, but is hardly sufficient for success. Lifting compatibility constraints is not the same thing as obtaining a successful monitoring system. A recipe for success goes further and entails in our context capturing the full and up-to-date range of indicators in a reliable fashion and validating such monitoring data either with automated or manual procedures. While manual validations relying on consistent data may substitute for the lack of automated checks, it turns out that the lack of socio-technical standards coupled with manual validations will spell out the failure of the monitoring system because of undetected errors, duplication of controls and frequent re-keying of data entries.

References

Basurto, X. & Speer, J. 2012. "Structuring the calibration of qualitative data as sets for qualitative comparative analysis," Field Methods (24:2), pp. 155-174.

Berg-Schlosser, D., De Meur, G, Rihoux, B. & Ragin, C. 2009. "Qualitative Comparative Analysis (QCA) as an Approach," in Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques, B. Rihoux, and C. Ragin (eds.), London: Sage, pp. 1-18.

Burton-Jones, A., McLean, E. R. & Monod, E. 2015. "Theoretical perspectives in IS research: from variance and process to conceptual latitude and conceptual fit," European Journal of Information Systems (24:6), pp. 664-679.

Cavaye, A. L. M. 1996. "Case study research: a multi-faceted research approach for IS," Information Systems Journal (6:3), pp. 227-242.

Chan, Y. E. 2000. "IT value: the great divide between qualitative and quantitative and individual and organizational measures," Journal of Management Information Systems (16:4), pp. 225-261.

Corbin, J. & Strauss A. 1990. "Grounded theory research: procedures, canons and evaluative criteria," Qualitative Sociology (13:1), pp. 3-19.

Cornford, T., Doukidis, G. & Forster, D. 1994. "Experience with Structure, Process and Outcome Framework for Evaluating an Information System," Omega International Journal of Management Science (22:5), pp. 491-504.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. 1989. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science (35:8), pp. 982- 1003.

Davis, G. B., Allen, L. S., Nickles, K. R., Chatterjee, S., Hartung, R. & Wu, Y. 1992. "Diagnosis of an information systems failure: a framework and interpretive process," Information and Management (23:5), pp. 293-318.

DeLone, W. H. & McLean, E. R. 1992. "Information Systems Success: The Quest for the Dependent Variable," Information Systems Research (3:1), pp. 60-95.

DeLone, W. H. & McLean, E. R. 2003. "The DeLone and McLean model of information systems success: a 10 year update," Journal of Management Information Systems (19:4), pp. 9-30.

Devaraj, S. & Kohli, R. 2003. "Performance impacts of information technology: Is actual usage the missing link?" Management Science (49:3), pp. 273-289.

Dobson, P., Myles, J. & Jackson, P 2007. "Making the case for critical realism: examining the implementation of automated performance management systems," Information Resources Management Journal (20:2), pp. 138-152.

Dunleavy, P., Margetts, H., Bastow, S. & Tinkler, J. 2006. Digital era governance: IT corporations, the state, and E-government, Oxford: Oxford University Press.

Eisenhardt, K. M. 1989. "Building theories from case study research," The Academy of Management Review (14:4), pp. 532-550.

El Sawy, O., Malhotra, A, Park, Y. & Pavlou, P. A. 2010. "Seeking the configurations of digital ecodynamics: it takes three to tango," Information Systems Research (21:4), pp. 835-848.

Fiss, P. C. 2011. "Building better causal theories: a fuzzy set approach to typologies in organization research," Academy of Management Journal (54:2), pp. 393- 420.

Gable, G. G. 1994. "Integrating case study and survey research methods: an example in information systems," European Journal of Information Systems (3:2), pp. 112-126.

George, A., and A. Bennett 2005. Case studies and theory development in the social sciences, Cambridge, Massachusetts: MIT Press.

Goertz, G. and Mahoney, J. 2012. A tale of two cultures: qualitative and quantitative research in the social sciences, Princeton, NJ: Princeton University Press.

Goodhue, D. H., Klein, B. D. & March, S. T. 2000. "User evaluations of IS as surrogates for objective performance," Information & Management (38:2), pp. 87-101.

Henfridsson, O., & Bygstad, B. 2013. "The generative mechanisms of digital infrastructure evolution," MIS Quarterly (37:3), pp. 907-931.

Kaplan, B. 2001. "Evaluating informatics applications- some alternative approaches: theory, social interactionism, and call for methodological pluralism," International Journal of Medical Informatics (64:1), pp. 39-56.

Kaplan, B. & Duchon D. 1988. "Combining qualitative and quantitative methods in information systems research: a case study," MIS Quarterly (12:4), pp. 571-586.

Kaplan, B. & Shaw, T. 2004. "Future directions in evaluation research: people, organizational, and social issues," Methods of Information in Medicine (43:3), pp. 215-231.

Klecun, E. & Cornford, T. 2005. "A critical approach to evaluation," European Journal of Information Systems (14:3), pp. 229-243.

Klecun, E., Lichtner, V., Cornford, T., & Petrakaki, D. 2014. "Evaluation as a Multi-Ontological Endeavour: A Case from the English National Program for IT in Healthcare," Journal of the Association for Information Systems (15:3), pp. 147-176.

Markus, L. M. & Robey, D. 1988. "Information technology and organizational change: causal structure in theory and research," Management Science (34:5), pp. 583-598.

Markus, M. L. & Robey, D. 2004. "Why Stuff Happens: Explaining the Unintended Consequences of Using Information Technology," in The Past and Future of Information Systems, M. T. Vendelo, and K. V. Andersen (eds.), Amsterdam, Netherlands: Elsevier Butterworth-Heinemann, pp. 61-93.

Mingers, J. 2001. "Combining IS research methods: towards a pluralist methodology," Information Systems Research (12:3), pp. 240- 259.

Pare´, G., Bourdeau, S.,Marsan, J., Nach, H. & Shuraida, S. 2008. "Re-examining the causal structure of information technology impact research," European Journal of Information Systems (17:4), pp. 403-416.

Pawson, R., and Tilley, N. 1997. Realistic Evaluation. London: Sage Publications.

Petter, S., DeLone, W. & McLean, E. R. 2012. The past, present and future of IS "success". Journal of the Association for Information Systems (13:5), pp. 341-362.

Polites, G., L., Roberts, N. & Tatcher, J. 2012. "Conceptualising models using multidimensional constructs: a review and guidelines for their use," European Journal of Information Systems (21:1), pp. 22-48.

Ragin, C. C. 1987. The comparative method: moving beyond qualitative and quantitative strategies. Berkeley: University of California Press.

Ragin, C. C. 2000. Fuzzy set social science. Chicago: The University of Chicago Press.

Ragin, C. C. 2008. Redesigning Social Enquiry. Chicago: The University of Chicago Press.

Rai, A., Patnayakuni, R. & Patnayakuni, N. 2006. "Firm performance impact of digitally enabled supply chain capabilities," MIS Quarterly (30:2), pp. 225-246.

Rivard, S.& Lapointe, L. 2012. "Information Technology Implementers' Responses to User Resistance: Nature and Effects," MIS Quarterly (36:3), pp. 897-920.

Schneider, C. Q., and Wagemann C. 2012. Set-Theoretic methods for the social sciences: a guide to qualitative comparative analysis. Cambridge: Cambridge University Press.

Seddon, P. B. 1997. "A respecification and extension of the DeLone and McLean model of IS success," Information Systems Research (8:3), pp. 240-253.

Sedera, D., Eden, R. & McLean, E. R. 2013. "Are we there yet? A step closer to theorizing information systems success," in Proceedings of the 29th International Conference on Information Systems, Baskerville, R. and Chau, M. (eds.), Milan, Italy. [WWW document] Available at: http://aisel.aisnet.org/icis2013/proceedings/ResearchMethods/7/ (Access date: 12/10/2015).

Serafeimidis, V. & Smithson, S. 2003. "Information systems evaluation as an organizational institution – experience from a case study," Information Systems Journal (13:3), pp. 251–274.

Strauss, A., and Corbin, J. 1998. Basics of Qualitative Research: Techniques and Procedures for developing Grounded Theory. Thousand Oaks, CA: Sage Publications.

Symons, V. 1991. "A review of information systems evaluation: content, context, process," European Journal of Information Systems (1:3), pp. 205-212.

Tödtling-Schönhofer, H., Naylon, I., Génin, H., Basile, I. & Savioli, B. 2011. Synthesis Report Metis GmbH in cooperation with Euréval. [WWW document] Available at: http://ec.europa.eu/social/BlobServlet?docId=7342&langId=en (Access date: 06/04/2016).

Van de Ven, A. 2007. Engaged Scholarship. A Guide for Organizational and Social Research Oxford: Oxford University Press.

Wixom, B. H. & Todd, P. A. 2005. "A Theoretical Integration of User Satisfaction and Technology Acceptance," Information Systems Research (16:1), pp. 85-102.

Yin R. K. 2014. Case Study Research: Design and Methods. Thousands Oaks, CA: Sage Publications.