

---

## Evaluation Criteria for Information Quality Research

---

### Philip Woodall\*

Institute for Manufacturing, Department of Engineering, University of Cambridge, 17 Charles Babbage Road, Cambridge, United Kingdom

E-mail: [phil.woodall@eng.cam.ac.uk](mailto:phil.woodall@eng.cam.ac.uk)

\*Corresponding author

### Alexander Borek

Gartner Deutschland GmbH, Lehrer-Wirth-Straße 2, 81829 München, Germany

E-mail: [alexander.borek@gartner.com](mailto:alexander.borek@gartner.com)

### Ajith Parlikad

Institute for Manufacturing, Department of Engineering, University of Cambridge, 17 Charles Babbage Road, Cambridge, United Kingdom

E-mail: [ajith.parlikad@eng.cam.ac.uk](mailto:ajith.parlikad@eng.cam.ac.uk)

**Abstract:** Evaluation of research artefacts (such as models, frameworks and methodologies) is essential to determine their quality and demonstrate worth. However, in the IQ research domain there is no existing standard set of criteria available for researchers to use to evaluate their IQ artefacts. This paper therefore describes our experience of selecting and synthesizing a set of evaluation criteria used in three related research areas of Information Systems (IS), Software Products (SP) and Conceptual Models (CM), and analysing their relevance to different types of IQ research artefact. We selected and used a subset of these criteria in an actual evaluation of an IQ artefact to test whether they provide any benefit over a standard evaluation. The results show that at least a subset of the criteria from the other domains of IS, SP and CM are relevant for IQ artefact evaluations, and the resulting set of criteria, most importantly, enabled a more rigorous and systematic selection of what to evaluate.

**Key Words:** Data quality, information quality, data quality evaluation, evaluation criteria, IQ artifacts, DQ artifacts, DQ evaluation, IQ evaluation

## 1 Introduction

Few people would denounce the importance of evaluation in the Information Quality (IQ) research domain. Without systematic and rigorous evaluations, the IQ community could suffer from a general lack of understanding, by both researchers and practitioners, of the overall quality of the research outputs. A previous review of IQ research recognized the importance of evaluation frameworks in IQ and called for future research to address this area (Madnick et al., 2009). Despite this, there is no collated set of criteria available to IQ researchers, which can be measured during an evaluation to predict the success of IQ artefacts (IQ artefacts being a convenient term to encompass various IQ research outputs, such as, models, frameworks, algorithms, methodologies etc.).

The aim of this paper is therefore to present a set of relevant evaluation criteria for IQ artefacts. The research was motivated by our own struggle with finding and selecting evaluation criteria so that we could perform a rigorous and systematic evaluation of an existing IQ artefact (see (Woodall, Borek and Parlikad, 2013)). Numerous criteria have been proposed in other research domains, but determining which of these we should select and which would be relevant to our artefact was a very time-consuming process.

### *Defining IQ artefacts and the approach for developing a set of evaluation criteria*

To define IQ artefacts more formally, we adapted the design science definition for IS artefacts (Hevner et al., 2004) and tailored this more specifically to IQ:

IQ artefacts are defined as:

- constructs (vocabulary and symbols),
- models (abstractions and representations),
- methods (algorithms and practices),
- instantiations (implemented and prototype systems)

which enable IQ researchers and practitioners to understand and address the problems inherent in successfully *improving information quality within organizations*.

A typical example of “constructs” are the use of IQ dimensions, which are a specific vocabulary that describe how IQ problems can be categorized and distinguished (see for example (Wang and Strong, 1996) and (Levitin and Redman, 1995)). For the second category, an example of a “model” is given by Folmer and Soest who propose a model for semantic information system standards (Folmer and Soest, 2012). “Methods” are common in the IQ domain and include the various methodologies for assessment and improvement of IQ (see (Batini et al., 2009) for examples), and more specific algorithms which, for example, can detect data errors (see for example (Meda, Sen and Bagchi, 2010)). Finally, “instantiations” include the various tools which can be used to correct quality problems (see (Friedman, 2012) for examples).

Rather than create a set of criteria from scratch, in this paper we argue that IQ artefacts are not completely distinct from research in other areas, and can therefore use the criteria from these areas. In particular, in the related areas of information systems (IS), software products (SP), and conceptual models (CM), there has been a vast amount of research

into how to properly evaluate artefacts and what set of quality criteria should be evaluated, and are vital, to determine the ultimate success of the artefact.

The question is whether the criteria from these other domains are suitable, for the evaluation of IQ artefacts. And hence our first research question centres on this point:

RQ 1: Are the evaluation criteria used for information systems, software products, and conceptual model quality suitable for the evaluation of IQ artefacts?

The IS, SP, and CM research areas were selected as candidates to provide suitable evaluation criteria because each area has some overlap with the IQ domain. The domains of IS and SP are clearly very closely related to IQ, with many of the IQ artefacts following the principles from software algorithm development and IS design and development e.g. (Dash and Singhanian, 2009; Li, 2009; Tremblay, Dutta and Vandermeer, 2010; Meda, Sen and Bagchi, 2010). Note that many of the IQ research artefacts are also very similar to conceptual models. For example, the well-known IQ methodology TDQM (Wang, 1998) includes a conceptual model—referred to as IP-MAP (Shankaranarayanan, Wang and Ziad, 2000)—of the information product that extends the Entity Relationship diagram used in software development.

Note also that the IS, SP and CM areas have had significant effort placed on developing evaluation criteria over many years and therefore provide established and reliable sources for evaluation criteria. Furthermore, even if these criteria are relevant to the IQ domain, different types of artefact may demand different criteria, and, therefore, we also consider the following question:

RQ 2: What criteria are suitable for evaluating each IQ artefact type (for example, constructs, models, methods and instantiations)?

Answering this second question helps to identify the set of criteria which researchers should consider measuring when they have a particular type of artefact. It also helps to determine whether it makes sense to try to find a single set of criteria from which all IQ artefacts can be evaluated; if each IQ artefact requires completely different criteria, then searching for the ‘general set’ is perhaps not the best approach.

To answer these research questions, we define what is meant by evaluation (see Section 2), then we collated an exhaustive set of evaluation criteria from the IS, SP and CM domains (see Section 3), synthesized them to remove duplicates and unify the terminology and identified the relevant artefact types for each criterion (see Section 4). To determine whether these criteria are suitable for IQ artefacts (RQ 1), we reviewed and classified various existing IQ research artefacts, and analysed to what extent the criteria have already been applied in the existing evaluations (see Section 5). Furthermore, we tested a subset of the synthesized evaluation criteria by using them in an actual evaluation of an IQ artefact (see Section 6). Finally, to answer RQ 2, we show how the criteria are spread between different artefact types by grouping the criteria according to artefact type (see Section 7). Finally, we present the conclusions in Section 8.

## **2 Definitions of evaluation**

There are numerous definitions of evaluation in different fields of research. For instance, in the field of Design Science, where it is the artefact that is the focus (Hevner et al., 2004), evaluation is defined as: “The process of determining how well the artefact performs” and this is measured by assessing a set of predetermined criteria (March and Smith, 1995). In IS, therefore, evaluation focuses on determining to what extent the IS is a success, with many different criteria used to measure this (DeLone and McLean, 1992). The evaluation may be conducted ex ante (prior to artefact construction – i.e. evaluation of the artefact from its specification) and/or ex post (after artefact construction) (Pries-Heje, Baskerville and Venable, 2008). An ex ante evaluation of an IS is typically done by organizations to decide whether or not to invest in the new IS (Walter and Spitta, 2004), whereas ex post evaluation helps to determine whether the chosen IS after implementation is a success or not (Farbey, Land and Targett, 1999).

Within the domain of software products, ISO 25040 defines evaluation as “the systematic determination of the extent to which an entity meets its specified criteria” (ISO/IEC 25040, 2011). This is intended for software products that need to be evaluated to decide whether the product meets the relevant quality characteristics. The ISO document also describes an evaluation process consisting of establishing the evaluation requirements, specifying the evaluation, designing, executing and finally concluding the evaluation. While the evaluation process is very important, this paper focuses only on the selection of evaluation criteria, which is one stage in the process—in the case of the ISO 25040, this process comes under “specify the evaluation” (ISO/IEC 25040, 2011).

The numerous definitions and types of evaluation all point to the common theme of assessing some artefact against predefined quality criteria to determine its overall quality. It was with this understanding of evaluation that we considered evaluation criteria and framed the scope of this research.

## **3 Building a set of evaluation criteria**

This section describes our efforts to collate a set of evaluation criteria from the related domains of IS, SP and CM. For each of the sets of evaluation criteria found, we discuss the development of the criteria and whether the proposed criteria are generally relevant for the evaluation of IQ artefacts.

### *Evaluation criteria for Information Systems*

In the area of information systems, evaluation work that considers criteria has focused on determining what criteria can be used to measure whether an IS is, or is not, a success. However, most of the evaluation research in IS, rather than focus particularly on developing criteria, focuses on general aspects of evaluation, such as: why the evaluation is being done? Who affects the evaluation? When is the evaluation taking place? How is the evaluation to be carried out? (Stockdale and Standing, 2006). General reviews of the vast topic of evaluation in information systems can be found in (Farbey, Land and Targett, 1999) and (Walter and Spitta, 2004).

Despite the variety of evaluation-related research in information systems, there is still no clear agreement on what precisely needs to be evaluated (Klecun and Cornford, 2005). Furthermore, measurement of success is difficult because in many cases the benefits are intangible and more than just cost-related (Irani, 2002; Symons, 1991). With regard to the choice of evaluation criteria for a particular study, (Klecun and Cornford, 2005) provide advice on the selection for a particular case, stating that it needs to be critically informed (i.e. cannot be divorced from the situation under study).

DeLone and McLean (1992) provided the key contribution regarding the development of a general set of evaluation criteria by developing a taxonomy containing the following six criteria (referred to as categories), which they used to classify the existing literature that measured IS success: system quality, information quality, use or intention to use, user satisfaction, individual impact and organizational impact.

This model has been tested and refined by numerous researchers (Seddon, 1997), (Grover, Seung Ryul Jeong and Segars, 1996) (see also the references in the ‘Model Issues’ section in (DeLone and McLean, 2003)) with the ultimate aim of adding greater clarity to the criteria and the relationships between them. The two key changes that impacted the criteria were the later introduction of the service quality criterion and the merging of individual impact and organizational impact into the single ‘net benefit’ criterion (DeLone and McLean, 2003). Table 1 shows the final list of criteria with the associated definitions. Note that no explicit definition was given in the more recent paper for the service quality or net benefits criteria.

<b>IS success measures</b>	<b>Measures previously referred to as</b>	<b>Definition</b>
System quality	-	Measures of the information processing system itself
Information quality	-	Measures of information system output
Service quality	-	No definition given
Use or intention to use	-	Recipient consumption of the output of an information system
User satisfaction	-	Recipient response to the use of the output of an information system
Net benefits	Individual impact	The effect of information on the behavior of the recipient
	Organizational impact	The effect of information on organizational performance

**Table 1: IS evaluation criteria**

*Relevance of the IS criteria to IQ artefacts*

System quality, use or intention to use, user satisfaction, and net benefits are all directly relevant to IQ artefacts. For instance, system quality relates to the internal quality of an IS itself and can be directly mapped to the quality of the internal properties of an IQ artefact

itself. For example, the system quality of an IQ related algorithm could refer to aspects such as the efficiency of the algorithm in terms of speed and space utilization.

In the information systems area, the information quality criterion relates to the quality of the output of the information system, and the service quality relates to the quality of the service offering of the IS. The analogue for IQ artefacts is therefore the quality of the output of the IQ artefact.

### *Evaluation criteria for Software Products*

A series of international standards have been developed to support the evaluation of software products. In particular, ISO25040 describes the recommended evaluation process, while ISO25010 provides quality models that can be used within the evaluation process (ISO/IEC 25010, 2011)(ISO/IEC 25040, 2011).

Two models contained within the ISO25010 standard are relevant to the IQ evaluations: the product model and the in-use model (see Table 2 and Table 3). The in-use model is intended to assist with the evaluation of the outcomes of interaction with the software product, and it contains the following criteria: effectiveness, efficiency, satisfaction, safety, and context comprehensiveness. The definitions for these criteria are shown in Table 3.

As a complement to the in-use model, the product model in ISO25010 attempts to capture the criteria that help to determine the inherent quality of the software product itself, including: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability (see Table 2). Therefore, the product model relates to the 'system quality' criterion in the IS success criteria because it aims to determine the quality of the internal properties of the system itself. In this case, the criteria of the software product model actually provide a more detailed view of the components of 'system quality' from the IS criteria.

<b>Product model evaluation criteria</b>	<b>Sub criteria</b>	<b>Definition</b>
Functional suitability		The degree to which the product provides functions that meet stated and implied needs when the product is used under specified conditions
	Completeness	The degree to which the set of functions covers all the specified tasks and user objectives
	Correctness	The degree to which the product provides the correct results with the needed degree of precision
	Appropriateness	The degree to which the functions are suitable for specified tasks and user objectives

Performance efficiency		The performance relative to the amount of resources used under stated conditions
Compatibility		The degree to which two or more systems or components can exchange information and/or perform their required functions while sharing the same hardware or software environment
Usability		The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use
Reliability		The degree to which a system or component performs specified functions under specified conditions for a specified period of time.
Security		The degree to which information and data are protected so that unauthorized persons or systems cannot read or modify them and authorized persons or systems are not denied access to them
Maintainability		The degree of effectiveness and efficiency with which the product can be modified
Portability		The degree to which a system or component can be effectively and efficiently transferred from one hardware, software or other operational or usage environment to another

**Table 2: Product model criteria for software products**

<b>In-use model evaluation criteria</b>	<b>Definition</b>
Effectiveness	The accuracy and completeness with which users achieve specified goals
Efficiency	The resources expended in relation to the accuracy and completeness with which users achieve goals
Satisfaction	The degree to which stakeholder needs are satisfied when a product is used in a specified context of use
Safety	The degree to which a product or system does not, under specified conditions, lead to a state in which human life, health, property, or the environment is endangered
Context comprehensiveness	The degree to which quality requirements are met in all the specified contexts of use

**Table 3: In-use model criteria for software products**

*Relevance of the SP criteria to IQ artefacts*

The main difference between software products and IQ artefacts is the level of concreteness of the final product. Many IQ artefacts are not actual implementations, but rather higher-level concepts and suggested procedures that may be implemented in different ways. In contrast, a software product is always a concrete product. For the in-use and product models, this distinction impacts the use of the criteria for IQ artefacts. For instance, the safety criterion is only relevant to actual implementations in a particular

context, because whether or not, for example, a human life is put in danger can only be determined in a particular context of use. The criteria from the SP domain therefore relate to an instantiation or the concrete use of an IQ artefact in a particular context.

### *Evaluation criteria for conceptual models*

Conceptual models, stated very generally, are “any collection of specification statements relevant to some problem” (Lindland, Sindre and Sølvsberg, 1994, p.42). Typical examples of conceptual models include Entity Relationship Diagrams (ERD) diagrams, data models, Unified Modelling Language (UML) diagrams, and these are often used in the development of databases, general software, and information systems. Determining the quality of conceptual models, which are typically developed early in the development process, provides an opportunity to identify and correct defects early in the development process, which is faster and cheaper compared to later in the process (Moody and Shanks, 2003). For this purpose, researchers have focused on developing frameworks for the evaluation of conceptual models, and two well-known frameworks are those developed by Lindland et al. (Lindland, Sindre and Sølvsberg, 1994) and (Wand and Weber, 1990).

Wand et al.’s framework is not as relevant as Lindland et al.’s because it focuses only on the process of conceptual modelling. Lindland et al. propose the following criteria for evaluating conceptual models: syntactic correctness, feasible validity, feasible completeness, and feasible comprehension. The use of the word “feasibility” reflects the trade-off between benefits and drawbacks of measuring the criteria completely and partially (Lindland, Sindre and Sølvsberg, 1994).

An important criterion that is not in this framework that features in the IS evaluation criteria, is the ‘net benefits’ (to the user) criterion. However, the Lindland et al. framework has since been extended in numerous ways, see (Nelson et al., 2012) and (Kitchenham et al., 2005). The later extension of the framework added the feasible test coverage and practical utility quality criteria (Kitchenham, Linkman and Linkman, 2005); the practical utility criteria clearly covers the net benefits gap.

Similar to the software product quality models, the Kitchenham et al. framework criteria (shown in Table 4) has two levels of criteria referred to as quality criteria and quality goals. For our purposes, we focus on the more detailed quality goals.

Quality criteria	Quality goals	Definition
Syntactic quality	Syntactic correctness	All the statements in the underlying conceptual model and its implementation are syntactically correct
Semantic quality	Feasible completeness	The model contains all statements about the domain that are relevant to the model
	Feasible validity	All the statements made in the model are correct
Pragmatic quality	Feasible comprehension	The model is adequately understood by its target audience
	Feasible understandability	As far as feasibly possible the model is presented in an understandable format



Test quality	Test coverage	The model has been adequately tested in terms of feasible test coverage
Value	Practical utility; generic model	The value of a generic model is the extent to which provides non-trivial insights into the phenomenon being modelled
	Practical utility; specialized model	The value of a specialized model is the extent to it improves the processes used in a user organization

**Table 4: The Kitchenham et al. framework for conceptual models**

*Relevance of the CM criteria to IQ artefacts*

The Kitchenham et al. framework is intended for expert-opinion based models and therefore fits very well to many IQ artefacts, which have been developed from expert opinion (see for example, the framework of data quality dimensions developed by Wang and Strong (1996) using expert opinion-based surveys, and an information quality management maturity model (Baškarada, 2008) built using a Delphi study). Furthermore, although the framework is intended for expert-opinion based models, the criteria are clearly general and could also be applicable to many other applications.

**4 Synthesized Quality Evaluation Criteria**

This section presents a list of collated criteria from the IS, SP and CM domains (see Table 5) where the overlapping criteria have been merged, the names and definitions have been adjusted so that they are consistent, and the type of artefact to which the criteria relate is shown. The artefact type has been recorded based on its originating domain (IS, SP or CM), hence one can be sure that the criteria are definitely relevant to this artefact type. However, this does not necessarily mean that the use of the criterion is limited to this artefact type: in fact, in the IQ domain, the criteria have been used for differing types (see Section 7). Even after combining the criteria, arguably, there may still be similarities between different criteria. While it would be ideal to have completely orthogonal criteria, in practice it is more important to have a comprehensive set to choose from, which may have slight overlaps, but can potentially cover all the particular IQ evaluation requirements.

The following procedure was used to synthesize the criteria and their definitions (the detailed results from each stage are available upon request to the authors):

1. Collate all the criteria from the three domains (CM, CS and IS) into a single list including their original definitions.
2. Identify the overlapping criteria in the list from step 1. For the overlapping criteria, a single name and definition of one criterion (or a combination) was selected to be the primary criterion for the final list of criteria. The type of artefact to which the criteria relate was also identified. For criteria from the CM domain the artefact type is “model” (as they refer to conceptual models) and for the SP and IS domains the artefact type is “instantiation” (as they refer to specific instances of a software product or information system).

3. Reword the definitions so that each is consistent. In particular: 1) Each definition now starts consistently with the following: “the degree to which...”. 2) Make any reference to the artefact explicit (e.g. “The degree to which the set of functions covers all the specified tasks and user objectives” changes to “The degree to which the set of functions in the artefact covers all the specified tasks and user objectives”).
4. Independently review the process for generating the final criteria and definitions. The second and third authors reviewed the outcomes of the above steps and documented their comments for each step.
5. Reconcile each point mentioned by the reviewers so that a consensus is reached regarding the changes made.

The resulting set of criteria generated by following this procedure is shown in Table 5. In the following paragraphs we comment on how the problems identified by the independent reviews were reconciled.

From the independent reviews, one of the main problem encountered and noted by one of the reviewers was the generality of the criterion “system quality” from the IS domain, which overlaps with many of the criteria from the CM and SP domains. Naturally, the criteria are not all pitched at the same level of granularity with some being very specific and others being more general. To reconcile this problem, the more specific criteria were favoured over the general ones, because having more criteria provides more options for an evaluator to choose from. The overlapping criteria are noted in the final column of Table 5, and “system quality” appears in this column to show the overlap with the more specific criteria.

One reviewer noted that there is a similarity between “intention to use” and “usability”: the degree to which an instantiation can be used by specified users affects the degree to which the recipient uses or intends to use the output of the instantiation; i.e. they cannot use the output if they cannot use the instantiation to generate the output. We chose to keep these separate because a user may not always use the instantiation itself, they may only use the output or they may use both. For example, the instantiation may provide its output automatically with no user interaction. Keeping these criteria separate therefore means that, depending on the case, an evaluation could choose to measure either one or both, depending on the nature of the artefact.

The synthesising of most criteria was straightforward as system quality was a commonly overlapping criterion and so only the definitions needed to be unified so that they were consistent. However, the information quality and service quality criteria and the value criterion merit further elaboration.

The information quality and service quality criteria have been combined into a single criterion referred to as output quality. It is more convenient to consider these criteria from the more abstract perspective of “output quality” because it refers to the output of any IQ artefact. If this is information, then it should be measured in terms of information quality, if it is a service, then it should be measured in terms of the quality of the service provided.

The value criterion is the result of merging four criteria: practical utility (generic model), practical utility (specialized model), individual impact, and organizational impact. The latter two were previously already combined into “net benefits” by IS researchers. To avoid the problem of differentiating between a generic and a specific model, we simply adopt the term “value”, which was originally proposed in the CM list of criteria (see Table 4).

Criteria	Artefact type	Definition	Overlapping criteria
Syntactic correctness	Models	The degree to which all the statements in the artefact are syntactically correct	System quality
Completeness	Models	The degree to which the artefact contains all statements about the domain that are relevant to the artefact	System quality, Feasible completeness
	Instantiations	The degree to which the set of functions in the artefact covers all the specified tasks and user objectives	System quality, Feasible completeness
Validity	Models	The degree to which all the statements made in the artefact are correct	feasible validity, Correctness, System quality
	Instantiations	The degree to which the artefact provides the correct results with the needed degree of precision	feasible validity, Correctness, System quality
Usability	Instantiations	The degree to which the artefact can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use	System quality
Comprehension	Models	The degree to which the artefact is adequately understood by its target audience	Feasible comprehension
Understandability	Models	The degree to which the artefact is presented in an understandable format	Feasible understandability
Test coverage	Models	The degree to which the artefact has been adequately tested	System quality
Satisfaction	Instantiations	The degree to which stakeholder needs are satisfied when the artefact is used in a specified context of use	User satisfaction
Value	Models	The degree to which the artefact improves the processes used in a user organization and/or provides non-trivial insights into the phenomenon being modelled	Practical utility (generic model) and (specialised model), Net benefits
	Instantiations	The degree to which the artefact positively affects the behaviour of the recipient and/or organizational performance	Practical utility (generic model) and (specialised

			model), Net benefits
Output quality	Instantiations	The degree to which the output of the artefact is satisfactory	Information quality and Service quality
Use or intention to use	Instantiations	The degree to which the recipient uses or intends to use the output of the artefact	
Functional suitability	Instantiations	The degree to which the artefact provides functions that meet stated and implied needs when the product is used under specified conditions	System quality
Appropriateness	Instantiations	The degree to which the functions in the artefact are suitable for specified tasks and user objectives	Appropriateness, System quality
Performance efficiency	Instantiations	The degree to which the performance of the artefact, relative to the amount of resources used under stated conditions, is satisfactory.	System quality
Compatibility	Instantiations	The degree to which two or more artefact can exchange information and/or perform their required functions while sharing the same hardware or software environment	System quality
Reliability	Instantiations	The degree to which the artefact performs specified functions under specified conditions for a specified period of time.	System quality
Security	Instantiations	The degree to which information and data related to the artefact are protected so that unauthorized persons or systems cannot read or modify them and authorized persons or systems are not denied access to them	System quality
Maintainability	Instantiations	The degree to which the artefact can be modified with effectiveness and efficiency.	System quality
Portability	Instantiations	The degree to which the artefact can be effectively and efficiently transferred from one hardware, software or other operational or usage environment to another.	System quality
Effectiveness	Instantiations	The degree to which users of the artefact achieve specified goals to a satisfactory level of accuracy and completeness.	
Efficiency	Instantiations	The degree to which resources are expended satisfactorily, in relation to the accuracy and completeness with which users achieve goals, when the artefact is used.	
Satisfaction	Instantiations	The degree to which stakeholder needs are satisfied when the artefact is used in a specified context of use.	
Safety	Instantiations	The degree to which the artefact does not, under specified conditions, lead to a state in which human life, health, property, or the environment is endangered	

Context comprehensiveness	Instantiations	The degree to which quality requirements are met in all the specified contexts of use of the artefact.	
---------------------------	----------------	--	--

**Table 5: Synthesised criteria**

### 5 Suitability of the Criteria to IQ Artefacts

To determine the suitability of the criteria to IQ artefacts, we reviewed a series of IQ artefact evaluations and recorded (in Table 6) retrospectively what criteria were considered during the evaluation. Clearly, most papers did not explicitly state the name of the criteria, and it was therefore necessary to interpret each evaluation to determine what criteria were actually covered—it is, hence, a retrospective assignment of criteria to the evaluations. Implicit in this assignment of criteria to the evaluations is the relevance of the criteria: each assignment essentially shows that the particular criterion is relevant to the IQ artefact through being measured as part of the evaluation. Determining what set of criteria have been applied therefore helps to answer question RQ 1, which focuses on whether the criteria from the other domains are relevant to IQ artefacts.

We reviewed all articles in JDIQ between the start of the journal and vol. 2, issue 2, 2011 (the papers that were available at the time of conducting this research) and selected those that performed an evaluation. JDIQ was chosen because it has published a good range of IQ artefacts and all with evaluations. Thirteen papers were selected, and for each artefact (one in each paper) we recorded the type of evaluation and the assigned relevant evaluation criteria (see Table 6). The type of the artefact is also recorded in the first column in square brackets as either:

- Constructs [C] (vocabulary and symbols),
- Models [Mo] (abstractions and representations),
- Methods [Me] (algorithms and practices),
- Instantiations [I] (implemented and prototype systems).

One problem encountered while assigning the criteria was distinguishing between the performance efficiency and efficiency criteria. The difference between these criteria appears to be too subtle for this purpose and therefore the term performance efficiency is used to cover both criteria.

Regarding the construction of artefacts, the design science literature shows the two core and distinct phases of “build” and “evaluate” (Hevner, 2007). The criteria for evaluation may also be used during the build phase to ensure that the resulting artefact is likely to be successful in the final evaluation. However, in order to use the criteria in this way it is recommended that different methods be employed: the quality framework by Kitchenham et al. (2005), specifies that methods introduced to achieve the criteria should be separated from the means to assess the criteria. For a particular artefact being evaluated, and for each criterion, one should specify:

1. The methods that will be used during the construction of the artefact to ensure that the quality criterion will be satisfied, and

2. The methods that will be used to check whether the criterion has been satisfied after the artefact has been developed.

While reviewing the JDIQ papers, we did not find any use of evaluation criteria other than at the evaluation phase. Hence, all of the criteria and methods used to measure these shown in Table 6 are for the evaluation phase only.

IQ artefact	Type of evaluation	Relevant criteria
Signature Embedding Framework (Ababneh, Ansari and Khokhar, 2009) [Mo]	Conducted efficiency tests of the framework	Performance efficiency
Methodology for Assessing Data Quality (Even and Shankaranarayanan, 2009) [Me]	Trialled the methodology on real-world data sets.	Test coverage
		Value
A Procedure to Develop Metrics for Currency (Heinrich, Klier and Kaiser, 2009) [Me]	Checked the approach adhered to the requirements from the literature.	Validity
	The utility of the approach is illustrated by using a real-world scenario and showing how the results can be linked to organizational profits.	Value Test coverage (very limited test cases)
A Bayesian Approach for Estimating and Replacing Missing Categorical Data (Li, 2009) [Me]	Conducted an experimental study, with a real data set, comparing the performance of the approach with two existing methods.	Performance efficiency
	The comparison against other widely used models gives some indication of value.	Value
Optimal Stopping: A Record-Linkage Approach (Moustakides and Verykios, 2009) [Me]	Compared simulation results with theoretical predictions to assess the performance of the approach.	Performance efficiency
	Theoretic justification of the properties of the approach using formal proofs.	Validity
A Bayesian Network Learning Algorithm (Sessions and Valtorta, 2009) [Me]	Conducted efficiency tests of the algorithm	Performance efficiency
	Obtained test results for different Bayesian Networks	Test coverage
Algorithm for Detecting Data Flow Errors in Workflows (Meda, Sen and Bagchi, 2010) [Me]	Theoretical approach to demonstrating the correctness of the algorithm.	Validity
	The algorithm was implemented in software and tested.	Test coverage
Algorithm to Discover Bias Patterns in Missing Data (Tremblay, Dutta and Vandermeer, 2010) [Me]	Conducted efficiency tests of the algorithm	Test coverage Performance efficiency
	Discussed applicability with a real example	Value
Mining Algorithm in Large Noisy Domains (Dash and Singhania, 2009) [Me]	Conducted efficiency tests of the algorithm for different cases	Test coverage Performance efficiency
	The comparison against other algorithms/methods gives some indication of value	Value
An Accuracy Metric (Fisher, Lauria and Matheus, 2009) [Mo]	Conducted efficiency tests of the metric	Test coverage Performance efficiency
	Described an example of how the method can be used to answer realistic questions	Value
A Model of DQ in Sensor Data Streaming Environments (Klein and Lehner, 2009) [Mo]	Discussed the utility of the approach based on an example scenario	Value

A Contingency Approach to Data Governance (Weber, Otto and Österle, 2009) [Me]	Reported the experience of applying the approach in an organization.	Value
		Usability
Graph-Based Name Disambiguation (Fan et al., 2011) [Mo] This framework contains: [Me] – valid path search algorithm	Conducted efficiency tests of components of the framework (this included testing runtime performance).	Test coverage
		Performance efficiency
	The utility of the framework is illustrated by using a real-world data set and comparison of against a state-of-the-art approach.	Value

**Table 6: Relevance of the criteria to IQ artefacts**

### *Relevance of the criteria from other domains*

The previously stated research question RQ 1 asks whether the criteria used in the other related domains are relevant to evaluate IQ artefacts. In fact, five of the criteria from the other domains have been applied to evaluate IQ artefacts, these are:

- Performance efficiency
- Test coverage
- Value
- Validity
- Usability

These criteria have hence been confirmed as being relevant for evaluations of IQ artefacts. Clearly it is only possible to demonstrate the relevance of the criteria, and the absence of the other criteria from this list does not mean that the others are not relevant for IQ artefacts—it may be that for other IQ evaluations not included in our review the criteria are in fact useful. However, further actual evaluations of IQ artefacts are needed to confirm the utility of the other criteria. One such evaluation is presented in the following section, and the criterion “completeness” was also determined to be applicable, hence the final list of criteria relevant for evaluations of IQ artefacts is:

- Performance efficiency
- Test coverage
- Value
- Validity
- Completeness
- Usability

## **6 Trialling the Evaluation Criteria**

In order to determine whether the synthesized criteria are relevant and usable for IQ artefacts, the synthesized list of criteria was used to support an actual evaluation of an IQ artefact called the Hybrid Approach (HA).

The HA was recently developed as a new approach to IQ assessment, and the aim of this approach is to show how a new and fully customized IQ Assessment Technique (AT) can be developed for any organization wanting to assess IQ. A number of ATs have already been proposed, and to develop the HA the existing ATs were identified and then divided

into their constituent activities. These activities were then analysed to understand the order in which they should be placed in and whether any activity is dependent on another activity. Finally, a four step process was developed that describes how to develop a new AT based on the existing activities and their ordering and dependency constraints.

The key components of the HA are therefore:

1. The model of the activities showing overlaps, ordering constraints and dependencies between activities
2. The procedure (method) for using the model to develop a new, customized AT

Based on these components, intuitively, we developed our own requirements for the HA evaluation:

1. Is the approach correct in itself (i.e. is the model correct)?
2. Can the HA procedure be followed (i.e. can the target users understand and execute the procedure)?
3. Does the HA provide any benefit to the target audience?

#### *The evaluation of the Hybrid Approach*

It was not feasible for us to measure all of the criteria from Table 5 because of time constraints, and therefore it was necessary to select the most important subset for the evaluation of the HA. This was done using the initial requirements for the evaluation (listed above), and the selected criteria are shown in Table 7. The first two criteria were selected to cover the evaluation of whether the model is correct in itself, usability was selected to determine whether the method can be followed by the target audience, and test coverage and value were selected to evaluate whether the HA could provide any benefit to the target audience. Notice that the artefact type has been selected for each criterion because there are two different artefact types in the HA; for example, usability refers primarily to the usability of the method (which of course includes some usage of the model).

Criteria	Artefact type	Definition
Completeness	Model	The degree to which the artefact contains all statements about the domain that are relevant to the artifact.
Validity	Model	The degree to which all the statements made in the artefact are correct
Usability	Method	The degree to which the artefact can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use
Test coverage	Method	The degree to which the artefact has been adequately tested
Value	Method	The degree to which the artefact improves the processes used in a user organization and/or provides non-trivial insights into the phenomenon being modelled

**Table 7: Criteria selected for the HA evaluation**



For the evaluation requirement of determining whether the HA is correct in itself, the benchmark list of criteria gives numerous criteria that refer to the inherent quality of the artefact. The list of criteria proved to be useful in helping to be systematic in the selection of what to evaluate while also helping to determine what parts of the HA are the most important to focus on.

The completeness and validity criteria were chosen because they clearly capture the major problems that the model of activities in the HA could suffer from. For instance, the inherent quality of the model depends on two key components: the set of existing ATs used for the approach and the set of activities extracted from these ATs. If either an AT or an activity is missed and not included in the model or an invalid AT or activity is included, then the model will be poor quality.

In order to ensure that the HA model would be functionally complete and valid we did not want to wait until the final evaluation of the HA (after its development) to check this. Therefore, our approach to the evaluation was started during the development of the HA model, and included deliberate attempts to ensure that we would achieve completeness and validity. This followed the advice from the quality framework by Kitchenham et al. (2005), noted before in Section 5, where for a particular artefact being evaluated, and for each criterion, one should specify:

1. The methods that will be used during the construction of the artefact to ensure that the quality criterion will be satisfied, and
2. The methods that will be used to check whether the criterion has been satisfied after the artefact has been developed.

We therefore followed this guidance in order to help ensure that we would achieve the criteria as well as checking later whether or not we did. Table 8 lists the methods that were applied for both attempting to ensure that the criteria would be met and the methods that were used to check the criteria.

<b>Criteria</b>	<b>Methods applied in an attempt to ensure the criteria are met</b>	<b>Methods applied to check achievement of the criteria</b>
Validity and Completeness	Obtained the existing ATs from the literature systematically and applied selection criteria to ensure validity	Conducted an additional independent literature search to check whether any new ATs could be found (completeness) and whether the existing ATs would be selected again (validity)
Validity and Completeness	Extracted the activities using a structured extraction procedure and used an independent review process to ensure validity	Used new ATs found from the additional literature search and applied an extraction procedure to check for existing and new activities in the new AT.  Implemented the activities in an organizational trial

Usability	Conference paper review and general comments from attending the conference	An independent data assessor used the approach with in an organization (supported only with the existing documentation) for a number of small DQ assessments.
Test coverage	Conference paper review and general comments from attending the conference	Conducted an organizational trial of the approach in a real scenario
Value	Constructed the HA using only existing ATs that have been implemented and successfully trialled	Conducted an organizational trial of the approach in a real scenario

**Table 8 Methods used in the HA evaluation for the criteria**

*Allocation of methods to ensure the criteria were met*

To achieve validity and completeness of the ATs, a systematic literature review was used to capture as many ATs as feasibly possible, and selection criteria were specified so that only valid and practically useful ATs would be selected using the review. To achieve validity and completeness of the activities, a structured extraction procedure was used to ensure that no activities would be missed from an AT and a peer review process was used to ensure the validity of the activities.

A conference paper, describing the current progress of the HA, was peer reviewed by three reviewers as part of the International Conference on Information Quality review procedure. The review process was used as a means to check the usability criteria. Although it was not possible to ensure that the reviewers gave specific attention to this, useful feedback was obtained in relation to how the approach could be better described and presented. In an attempt give the HA some level of value, the HA was constructed by selecting only existing ATs that had been subject to a previous trial and implementation.

*Allocation of methods to check achievement of the criteria*

To perform a final assessment of the extent to which the quality criteria had/had not been achieved for the HA, a series of methods were carried out for each criterion.

To assess validity and completeness of the list of ATs, an additional literature search was carried out by an independent researcher to check for any new studies. Checking whether the existing ATs would be selected again helped to assess validity and checking for new studies helped to determine completeness. This literature review did find two new ATs, and these were used to confirm the validity and completeness of the existing set of activities. This was done by extracting the activities from the new ATs and checking for existing (validity) and new (completeness) activities in the new AT.

A series of smaller trials of the HA were carried out by an independent data assessor to evaluate the usability of the approach. This tested whether, using the documentation of the approach, the data assessor could carry out the approach successfully without assistance from the developers of the approach.

For the value criterion, the approach was trialled in an organization in a real scenario and analysed to check whether the approach provided any benefit to the organization after its use.

During the evaluation of the HA, the extra criteria “future resilience” was also considered; although, this was added later in the evaluation and was therefore only used as a method to check achievement of the criteria. Future resilience is defined with reference to the other quality criteria, and hence has not been included in the synthesized list. It is defined as “the other quality criteria remain stable or improve as the approach is used”. This ensures that the artefact does not pass all criteria at the time of the evaluation, but fails the test of time as it is used in the future.

### *Discussion of the findings from applying the criteria to the evaluation*

The key benefits observed while using the list of criteria (in Table 5) for the evaluation of the Hybrid Approach are that the criteria:

1. Enabled a more rigorous and systematic selection of what to evaluate,
2. Took the focus off having to develop evaluation criteria and enabled the efforts to be placed on how to perform the evaluation itself.

During the evaluation of the HA, each criterion was reviewed for relevance to the HA as well as its level of importance to the evaluation. This provided the reassurance that we had not overlooked an important part of the evaluation. Furthermore, by comparing the criteria against one another, it was easier to identify the criteria that were most important to measure for our particular evaluation.

Regarding the second benefit, not only did the criteria help to focus attention on the evaluation itself, but also they provided a constant reference point that indicated precisely what we needed to measure; developing a method for how to evaluate validity, for example, took multiple iterations and it was useful to have the overall goal for what needed to be measured (i.e. the definition of validity) constant and explicitly recorded.

## **7 Evaluation Criteria for each Artefact Type**

In order to answer the second research question (What criteria are suitable for evaluating each IQ artefact type?), based on the results from Table 6, we grouped the criteria according to the artefact types to which they have been applied in the IQ domain. Table 9 shows these results, which includes the IQ artefact type, the number of artefacts of that type (artefact type frequency), the evaluation criteria measured for that artefact type, the number of times the criterion was measured for the artefact type (criterion frequency), and examples of the ways in which the criteria were measured in the IQ domain. These were extracted from the review of JDIQ evaluations (see Table 6) and the Hybrid Approach evaluation; the evaluation criteria that were used for the evaluation of the Hybrid Approach are shown separately in round brackets.

IQ artefact type	Artefact type frequency	Evaluation criterion measured	Criterion frequency	Example measurement options
Models	4 (5)	Performance efficiency	3	Implement parts of the model and conduct efficiency tests.  Use a simulation to measure the results produced by the model in different cases.
		Test coverage	2	Run an experiment to observe the output of the model in different test cases.
		Value	3	Describe examples of how the proposed model could be used to answer realistic questions or by presenting a realistic example scenario.
		(Validity)	(1)	Check if the proposed model adheres to the requirements in the literature.
		(Completeness)	(1)	Check whether after repeating the literature review, nothing more significant is found.
Methods	9 (10)	Test coverage	6 (7)	Implement the method (for example in software) and check the output of the method in different test cases.
		Value	6 (7)	Use an experimental study with a real dataset to measure and compare the results against existing methods.
		Validity	3	Use theoretic approaches, such as formal proofs.
		Performance efficiency	5	Use an experimental study with a real data set to compare the performance of the method with other existing methods.  Use a simulation and compare the results with theoretical predictions to assess performance.
		Usability	1 (2)	Report the experience of trialling the method in an organization (an independent assessor can be used to reduce bias).

**Table 9: Usage of the evaluation criteria for different types of IQ artefact**

For the models artefact type, performance efficiency was measured by conducting efficiency tests which require implementing parts of the model, selecting metrics and measuring the metrics. Such metrics will be specific to the application, for example, the Peak signal-to-noise ratio (PSNR) metric used in (Ababneh, Ansari and Khokhar, 2009) is specific to image processing. Another case executed a simulation and measured the results produced by the model in different cases in the simulation. Hence, this also formed a measurement of the test coverage criterion as well as the performance efficiency criterion. In all three cases, the value criterion was discussed, rather than measured, by describing examples of how the proposed model could be used to answer realistic questions or by presenting a realistic example scenario.

For validity and completeness a literature review was used to check if the proposed model adhered to the requirements in the literature, and in the other case, the review checked whether after repeating the literature review, the same conclusions would arise. This latter literature review was also used to measure the related criterion of completeness by determining if any additional findings would be observed that had not been seen before.

For the methods artefact type, the validity criterion was checked using theoretic approaches, such as formal proofs. Performance efficiency was measured for the methods type in the same ways as for the models artefact type: using simulations and efficiency tests. In one case, performance efficiency was measured using an experimental study with a real dataset and comparing the results against existing methods (Li, 2009), and hence this also formed an evaluation of the value criterion. Value was often linked with measuring test coverage, which was measured usually by implementing the method (or the key sub parts of the method) in, for example, software or as a simulation and applying typical software tests. The value criterion was also evaluated if any part of the testing used real world datasets or scenarios, or was done as a trial in an organization. The usability of a method was also evaluated by reporting the experience of trialling the method in an organization, in one case, an independent assessor was used to ensure that the knowledge of the researcher who produced the method would not bias the results.

In summary, the criteria with associated artefact types (including the artefact types from the original frameworks of IS, SP and CM) which have been identified as being relevant to the IQ domain are:

- Performance efficiency (models, methods, instantiations)
- Test coverage (models, methods)
- Value (models, methods, instantiations)
- Validity (models, instantiations)
- Completeness (models, instantiations)
- Usability (methods, instantiations)

## 8 Conclusions and Future Work

Evaluation is a critical endeavour for the IQ research field and is needed to help demonstrate the worth of research artefacts such as frameworks, models and algorithms.

This paper describes our attempt to determine a suitable set of criteria that IQ artefacts can be evaluated against by synthesizing the existing criteria from other related domains. The first research question asked whether the criteria from the other domains of information systems, software products, and conceptual model quality are suitable for the evaluation of IQ artefacts. The retrospective review of IQ evaluations (Section 5) shows that there is an overlap with the criteria, and it was possible to determine, from the existing IQ evaluations, which criteria were used. Furthermore, the evaluation of the HA confirmed that the completeness criterion is also relevant. Overall, the following six criteria were noted as being relevant to the IQ domain: performance efficiency, test coverage, value, validity, completeness, and usability.

This final list of criteria provides a sample set of criteria for future evaluations of IQ artefacts to consider measuring. The final choice will clearly depend on the specific artefact and timescale for the evaluation, and hence a subset or superset (selected from the entire list in Table 5) may be considered.

The second research question focussed on determining which criteria are suitable for evaluating each IQ artefact type (for example, constructs, models, methods and instantiations). Only models and methods were observed in the IQ literature and the HA evaluation also focussed on a model and a method. Hence, future work could investigate what criteria are suitable for constructs and instantiations in particular. Note that although instantiations have been included in the list in Section 7 (because of the classification in the original frameworks from where the criteria originate) no actual IQ instantiations were observed in the literature.

The actual use of a subset of the criteria for an evaluation of an IQ artefact in Section 6 demonstrated how the benchmark set of criteria can be used to select the relevant criteria for a particular evaluation. It also gives an indication that the criteria are useful to help guide what needs to be measured during an evaluation.

Overall, it is hoped that further development of a benchmark set of evaluation criteria for the IQ domain will enable future evaluations to clearly explicate what is being measured. It can also help evaluations focus on the most important criteria while providing a consistent mechanism for comparison across different IQ evaluations.

## **Acknowledgements**

The authors gratefully acknowledge the financial support provided by EPSRC (grant EP/G038171/1).

## **References**

- Ababneh, S., Ansari, R. and Khokhar, A., (2009). Compensated Signature Embedding for Multimedia Content Authentication. *Journal of Data and Information Quality (JDIQ)*, 1 (3), Article 17.
- Başkarada, S., (2008). *IQM-CMM: Information Quality Management Capability Maturity Model*. Ph.D. thesis. Australia: University of South Australia.

- Batini, C. et al., (2009). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys*, **41** (3), pp.1–52.
- Dash, M. and Singhania, A., (2009). Mining in Large Noisy Domains. *Journal of Data and Information Quality (JDIQ)*, **1** (2), Article 8.
- DeLone, W.H. and McLean, E.R., (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, **3** (1), pp.60–95.
- DeLone, W.H. and McLean, E.R., (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, **19** (4), pp.9–30.
- Even, A. and Shankaranarayanan, G., (2009). Dual Assessment of Data Quality in Customer Databases. *Journal of Data and Information Quality (JDIQ)*, **1** (3), Article 15.
- Fan, X. et al., (2011). On Graph-Based Name Disambiguation. *Journal of Data and Information Quality (JDIQ)*, **2** (2), Article 10.
- Farbey, B., Land, F. and Targett, D., (1999). Moving Is Evaluation Forward: Learning Themes and Research Issues. *The Journal of Strategic Information Systems*, **8** (2), pp.189–207.
- Fisher, C.W., Lauria, E.J. and Matheus, C.C., (2009). An Accuracy Metric: Percentages, Randomness, and Probabilities. *Journal of Data and Information Quality (JDIQ)*, **1** (3), pp.1–21.
- Folmer, E. and Soest, J.V., (2012). Towards a quality model for semantic IS standards. *International Journal of Information Quality*, **3** (1), pp.28–48.
- Friedman, T., (2012). *Magic Quadrant for Data Quality Tools*, Gartner report.
- Grover, V., Seung Ryul Jeong and Segars, A.H., (1996). Information Systems Effectiveness: The Construct Space and Patters of Application. *Information and Management*, **31** (4), pp.177–191.
- Heinrich, B., Klier, M. and Kaiser, M., (2009). A Procedure to Develop Metrics for Currency and Its Application in CRM. *Journal of Data and Information Quality*, **1** (1), Article 5.
- Hevner, A.R. et al., (2004). Design Science in Information Systems Research. *MIS Quarterly*, **28** (1), pp.75–105.
- Hevner, A.R., (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, **19** (2), article 4.
- Irani, Z., (2002). Information Systems Evaluation: Navigating Through the Problem Domain. *Information and Management*, **40** (1), pp.11–24.
- ISO/IEC 25010, (2011). ISO/IEC 25010:2011 Systems and Software Engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE).
- ISO/IEC 25040, (2011). ISO/IEC 25040:2011 Systems and Software Engineering -- Systems and

Software Quality Requirements and Evaluation (SQuaRE).

- Kitchenham, B., Linkman, S. and Linkman, S., (2005). Experiences of Using an Evaluation Framework. *Information and Software Technology*, **47**, pp.761–774.
- Kitchenham, B.A. et al., (2005). A Framework for Evaluating a Software Bidding Model. *Information and Software Technology*, **47** (11), pp.747–760.
- Klecun, E. and Cornford, T., (2005). A Critical Approach to Evaluation. *European Journal of Information Systems*, **14** (3), pp.229–243.
- Klein, A. and Lehner, W., (2009). Representing Data Quality in Sensor Data Streaming Environments. *Journal of Data and Information Quality (JDIQ)*, **1** (2), Article 10.
- Levitin, A. and Redman, T., (1995). Quality dimensions of a conceptual view. *Information Processing & Management*, **31** (1), pp.81–88.
- Li, X.B., (2009). A Bayesian Approach for Estimating and Replacing Missing Categorical Data. *Journal of Data and Information Quality (JDIQ)*, **1** (1), Article 3.
- Lindland, O.I., Sindre, G. and Sølvsberg, A., (1994). Understanding Quality in Conceptual Modeling. *IEEE Software*, **11** (2), pp.42–49.
- Madnick, S.E. et al., (2009). Overview and Framework for Data and Information Quality Research. *Journal of Data and Information Quality*, **1** (1), Article 2.
- March, S.T. and Smith, G.F., (1995). Design and Natural Science Research on Information Technology. *Decision Support Systems*, **15** (4), pp.251–266.
- Meda, H.S., Sen, A.K. and Bagchi, A., (2010). On Detecting Data Flow Errors in Workflows. *Journal of Data and Information Quality (JDIQ)*, **2** (1), pp.1–31.
- Moody, D. and Shanks, G., (2003). Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework. *Information Systems*, **28** (6), pp.619–650.
- Moustakides, G.V. and Verykios, V.S., (2009). Optimal Stopping: A Record-Linkage Approach. *Journal of Data and Information Quality (JDIQ)*, **1** (2), Article 9.
- Nelson, H.J. et al., (2012). A Conceptual Modeling Quality Framework. *Software Quality Journal*, pp.1–28.
- Pries-Heje, J., Baskerville, R. and Venable, J., (2008). Strategies for Design Science Research Evaluation. In *Proceedings of the 16th European Conference on Information Systems*. pp. 255–266.
- 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.
- Sessions, V. and Valtorta, M., (2009). Towards a Method for Data Accuracy Assessment Utilizing a Bayesian Network Learning Algorithm. *Journal of Data and Information Quality (JDIQ)*, **1** (3), Article 14.



- Shankaranarayanan, G., Wang, R.Y. and Ziad, M., (2000). IP-MAP: Representing the Manufacture of an Information Product. In *Proceedings of the 2000 Conference on Information Quality*. pp. 1–16.
- Stockdale, R. and Standing, C., (2006). An Interpretive Approach to Evaluating Information Systems: A Content, Context, Process Framework. *European Journal of Operational Research*, **173** (3), pp.1090–1102.
- Symons, V.J., (1991). A Review of Information Systems Evaluation: Content, Context and Process. *European Journal of Information Systems*, **1** (3), pp.205–212.
- Tremblay, M.C., Dutta, K. and Vandermeer, D., (2010). Using Data Mining Techniques to Discover Bias Patterns in Missing Data. *Journal of Data and Information Quality (JDIQ)*, **2** (1), Article 2.
- Walter, D.-K. /MAIB S.G. and Spitta, P.D.-I.T., (2004). Approaches to the Ex-ante Evaluation of Investments into Information Systems. *Wirtschaftsinformatik*, **46** (3), pp.171–180.
- Wand, Y. and Weber, R., (1990). An Ontological Model of an Information System. *Software Engineering, IEEE Transactions on*, **16** (11), pp.1282–1292.
- Wang, R.Y., (1998). A Product Perspective on Total Data Quality Management. *Communications of the ACM*, **41** (2), pp.58–65.
- Wang, R.Y. and Strong, D.M., (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, **12** (4), pp.5–34.
- Weber, K., Otto, B. and Österle, H., (2009). One Size Does Not Fit All—A Contingency Approach to Data Governance. *Journal of Data and Information Quality (JDIQ)*, **1** (1), Article 4.
- Woodall, P., Borek, A. and Parlikad, A.K., (2013). Data quality assessment: The Hybrid Approach. *Information & Management*, **50** (7), pp.369–382.