AN EXPLORATORY STUDY ON KNOWLEDGE QUALITY THROUGH THE PEER REVIEW PROCESS FOR RESEARCH PUBLICATIONS

FARZAD SABETZADEH

Ph.D

The Hong Kong Polytechnic University

2016

THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

An Exploratory Study on Knowledge Quality through the Peer Review Process for Research Publications

Farzad Sabetzadeh

A Thesis Submitted in Partial Fulfilment of the Requirements

for the Degree of Doctor of Philosophy

November 2015

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

___________________ (Signed)

__________Farzad Sabetzadeh__________ (Name of Student)

ABSTRACT

In the fast growing world of science, knowledge is the building block of such an expansion. With the various sources of obtaining information over the World Wide Web, there is always concern about the quality of knowledge that is being shared, used or applied. There are many thousands of scientific papers that are reviewed, submitted and published and immediately become accessible over the internet. While much of the material on the web may look like scattered pieces of information, when these pieces are combined, structured and properly delivered, they become prime sources of knowledge on a very large scale. Data and information quality has been much talked about but when it comes to knowledge, explanation of quality is not an easy task due to the subjectivity of the quality and the context in which the knowledge is generated, disseminated or adopted. Peer-reviewing in the scientific publication context is one of the areas that gives quality assessment feedback for knowledge generation, dissemination and adoption as the building blocks of the knowledge value chain.

In this study, we first explore the fundamental definitions related to quality and knowledge to get a deeper insight on how to define knowledge quality in such a way it can cater for various contexts with different characteristics. These fundamental concepts are discussed based on the long-established epistemological terms being used over decades of scholarly endeavours. Such definitions are then expanded into the assessment of quality of knowledge in the context of the peer-review process in order to propose a working definition of knowledge quality in scientific contexts that is in concordance with all quality aspects described in the literature review. Scientific publications, as the ultimate output of the peer-review process, are good representatives for proper demonstration of knowledge quality that emerges in assessing research manuscripts. Various dimensions of knowledge quality are also explored and discussed throughout the peer-review experiment in this study.

The second part of this study explores the biases that affect knowledge quality assessments in peer-review process. These biases will be investigated through the knowledge quality attributes that are being used and understood by the reviewers in the peer-review process. These quality attributes are then scrutinized in connection with the knowledge resources for both content and schema (format) of the manuscripts. Furthermore, the decision patterns resulting from these knowledge quality attributes will be examined both quantitatively and qualitatively in order to identify the reviewer's preferences and priorities that may lead to peer-review decisions.

The last part of this study investigates the inter-subjectivity and agreement level among reviewers. It also examines the scaling impact for convergence or divergence of peer-review decisions when reviewers are randomly divided into groups or when the reviewers are increased or decreased from the decision panel (assuming the final decision is the one voted by majority). Such validations also demonstrate the importance of reliability in knowledge quality attributes as well as the implications in the generalization of these metrics within the same context. The qualitative analysis of the study will corroborate the extent of generalizability achieved from such quantitative measures.

This significance of this study is in portraying knowledge quality assessment in a tangible context like the peer-review process and the importance of proper understanding of quality in the context of knowledge in evaluating scientific works.

This exploratory study gives a better insight on how the knowledge quality is defined and assessed in a real world case study like peer-reviewing and how it can be systematically improved by analysis of different dimensions of knowledge quality for each context.

PUBLICATIONS ARISING FROM THE THESIS

Journal papers

1. Farzad Sabetzadeh , Eric Tsui , (2015) "An effective knowledge quality framework based on knowledge resources interdependencies", VINE, Vol. 45 Issue: 3, pp.360 – 375 , http://dx.doi.org/10.1108/VINE-07-2014-0048

Conference papers

1- Sabetzadeh, Farzad.; Tsui, E.; Lee, W.B., "Enhancing knowledge quality via a semantic-oriented framework for a social knowledge cloud,", 2014 5th IEEE International Conference in Software Engineering and Service Science (ICSESS), vol., no., pp.153-156, 27-29 June 2014 doi: 10.1109/ICSESS.2014.6933534

2- Sabetzadeh, F.; Tsui, E.; Lee, W.B., "Assessment of uncertainty in the quality of knowledge in the research publication review process," in Fuzzy Systems and Knowledge Discovery (FSKD), 2013 10th International Conference on, vol., no., pp.946-950, 23-25 July 2013 doi: 10.1109/FSKD.2013.6816331

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to my chief supervisor, Prof. Eric Tsui, whose his guidance during the development of my research topic helped me to be able to formulate my ideas into research problems. He has also spent much of his valuable time to review and comment on every stage of my research work, including this thesis, for integrity of the content and presentation of ideas.

I would also like to thank Prof W.B. Lee as my co-supervisor who has always provided me with valuable insight and research ideas that has helped me to have a better comprehension about research problems in my field of study. Many of his suggestions are embedded in the structure of this thesis.

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1 : INTRODUCTION

This chapter provides a summary of topics that will be investigated throughout this study. It discusses the importance of properly defining knowledge quality and explains such importance in decision making in the peer-review process, which are based on the reviewer's quality perception of a particular research work. By introducing the role of knowledge quality in these decisions, the reasons behind these decisions are explored to better identify the research questions and objectives that are proposed and examined throughout this study. The last section of this chapter shows the organization of this thesis.

1.1 Background of Study

Every day, there are many research works that are submitted to journals or scientific research panels to be peer-reviewed and assessed based on the quality of knowledge they deliver if they are published. The decision on accepting, revising or rejecting these research works plays an important role in the contribution to the world of science. While the purpose of the peer-review process is to monitor and maintain the quality of knowledge in scientific research publications (Tom Jefferson, Wager, & Davidoff, 2002), such quality standards significantly rely on the human judgment which is prone to subjectivity in decision-making (Langfeldt, 2006). This subjectivity in knowledge quality assessment, resulting in various peer-review decisions, can be observed from two different angels. The first one originates from the reviewer itself, rooted in his or her beliefs and characteristics about the proposed ideas, methodologies, applications etc. that are reflected and reported in a research study. The second one comes from deviation from the reviewer's proxy measures and norms that are taken into consideration for a decision in justifying a research work based on assumptions, methodologies, conclusions etc. that are proposed (C. J. Lee, Sugimoto, Zhang, & Cronin, 2013). These two extrinsic (reviewer's characteristics) and intrinsic (reviewer's scientific perception) elements in the peer-review process may appear in different degrees and formats during the peer-review process, shaping the final decision on the research work. However, the decision output from the peer-review process is based on the level of agreement reached from number of individuals. Thus, the peer-review process normally adopts the inter-subjectivity approach, where, subjective personal decisions should aim to generate universal views collectively (Bornmann, 2008; Lamont & Huutoniemi, 2011)

One of the ultimate goals of the evaluation of potential research publications is assessing the quality of knowledge that resides in those scientific manuscripts and what those research works can contribute to the world of science. Peer-reviewing is one the most commonly used processes for many years to assess the quality of knowledge achieved from these research works (Blackburn & Hakel, 2006). Indeed, knowledge quality can be best reflected in the assessment of published scientific papers through peer-reviewing, because scientific research publications do contribute to knowledge generation, dissemination and adoption as the major elements of evaluation in the knowledge value chain. While there have been various studies in the last 20 years on data and information quality that have resulted in the established terms and discipline in these two areas (English, 2009; Eppler, 2006; Redman & Godfrey, 1996), the knowledge quality concept is still vague and not well-defined (Newton, 2010; Poston & Speier, 2005; Rao & Osei-Bryson, 2007). Up to now, many of the research works that has been conducted in recent years on elucidating a knowledge quality definition, the relevant studies all tried to define a quality for knowledge from different perspectives (Rao & Osei-Bryson, 2007; Yoo, Vonderembse, & Ragu-Nathan, 2010). The absence of a proper definition for the quality of knowledge has also added to the subjectivity of knowledge quality assessment and decision outcomes. In the case of peer-reviewing, such ambiguity in defining knowledge quality increases decision biases arising from the intrinsic factors in the peer review process (reviewer's scientific perception). This is because, the scientific proxy measures used by peer reviewers to assess the quality of a research work may deviate significantly even if knowledge quality assessment is done within the same scientific discipline. Therefore, there is clearly a need to come up with a working definition of knowledge quality that can encompass all the quality aspects and can be initiated in a scientific context.

While peer-reviewing has been widely accepted as an effective way of assessing the quality of published work over the past decades, with the easy accessibility of any individual to the worldwide web, there is scepticism about the suitability of the peer-review process (or blind peer-review) as the most valid yardstick in assessing the quality of knowledge in published research works in the connected world today (Hardaway & Scamell, 2012; Smith, 2006). Nowadays, while some scholars insist on the credibility of blind peer-review that takes place in the majority of the scientific publications (Nicholas et al., 2015), some other scholars are either sceptic about the peer-review efficiency and effectiveness in evaluating quality of knowledge or they suggest alternative ways in order to improve the peer-review process (Resnik & Elmore, 2015)

With the rapid growth of knowledge and the wide and instant accessibility to any published scientific material over the world wide web, the suitability and purposefulness of the knowledge quality measures adopted and used by reviewers and the effectiveness and efficiency of the peer-reviewing method can have a significant impact on the future of the science world as a whole.

1.2 Research Motivation and Problem statement

As briefly implied in the previous section, there are three major challenges that need to be taken into consideration in the peer review process. Firstly, **(I)** Knowledge quality assessment decisions come from reviewers who are ultimately human beings with expertise or sometimes with little competent knowledge in the scientific domain of the content they are reviewing. The subjectivity in the decision that comes from a reviewer, either as a result of the personal characteristics of the reviewer or the preferences in the quality measures being taken by a reviewer to assess a research work has led to a lot of controversies (C. J. Lee et al., 2013; Resnik & Elmore, 2015; Smith, 2006). Secondly, **(II)** with no unilateral expression for defining the quality of knowledge in the peer-reviewing process, many of the knowledge quality assessment criteria might be chosen arbitrarily based on the common norms regarded to be important by a publisher, a journal or a panel (Goodman, Berlin, Fletcher, & Fletcher, 1994; Tom Jefferson et al., 2002; Weber, Katz, Waeckerle, & Callaham, 2002). Thirdly, **(III)** with the accessibility of the huge amount of scientific publications over the worldwide web, there are debates on whether traditional peer reviewing in the form of blind peer-review with a few reviewers is as efficient and effective as opening it up to the world as an open peer-review (Hardaway & Scamell, 2012; Suárez, Bernhard, & Dellavalle, 2012). In other words, whether having more reviewers and hence, more decision makers will give any better estimates about the decision output on the assessment of quality of knowledge for a manuscript.

Resnik & Elmore (2015) have summarized many of the studies conducted in all these three areas over the past 20 years. As many of these studies suggest, some of them do support the traditional blind-peer reviewing as the basic (and reliable) mechanism to maintain the integrity of knowledge in every discipline. On the other hand, many research studies question the fairness of the reviewers (Challenge I) and the validity of the assessment criteria used to evaluate the quality of knowledge for a manuscript (Challenge 2). Finally, as has been mentioned in the report of Resnik & Elmore (2015), very few studies have tried to trace the result of decisions that came from a peer-review on a longer term, known as predictive validity. The number of citations for a published research work may or may not show the agreement by many readers of that reviewed content. Nevertheless, there is a dispute as to whether open peer review can be a more promising evaluation method to the traditional peer review in a long run (Goodman et al., 1994) or there isn't such a big difference in the knowledge quality decision output by scaling the size of the decision (Levis, Leentjens, Levenson, Lumley, & Thombs, 2015; Nicholas et al., 2015) (Challenge III).

The aim of this study is to observe these three challenges not only from the statistical point of view that has been explored and discussed in many previous studies on peer review, but to see it in action, how the peer review process is performing in a real life scenario of knowledge quality assessment. This study initially explores deeply into the concepts of the quality of knowledge that has been barely explored in previous studies. It continues, by setting up a real life experiment based on the commonly used quality assessment measures that are used by the reviewers for assessment, and compares the biases and the knowledge quality attributes used in those quality assessment criteria in both quantitative and qualitative ways. These assessment criteria can be compared against the definition of quality criteria for knowledge to show a better picture of the elements that can influence the reviewer's decisions that might have originated from any of these three challenges.

1.3 Research Questions and Objectives

On the basis of the challenges that have been discussed so far, this research study pursues three major objectives. The first objective is **(A)** to give a better definition and understanding about knowledge quality and how it can be best reflected in the knowledge quality assessment of research publications. As there have been very few studies on this topic, such description for knowledge quality can show the significance and contributory role in understanding what should be sought and explored in the peer-review process and how it can be improved. This objective will be the foundation of the conclusion part that compares the real-life peer-review experiment outcomes with what has been defined as the target in definition of quality of knowledge in peer-review. The second objective is **(B)** to explore the decision patterns, in both quantitative and qualitative ways, with different sets of knowledge quality attributes to better reflect biases based on the background and preferences of the reviewers in making a decision on publication. This objective aims to analyse these decisions based on what has been reflected in the form of the reviewer's content and schema (format) preferences for a scientific manuscript. The third and the last objective is **(C)** to compare the effect of changing the number of reviewers and random distribution of reviewers assigned to assess the quality of manuscript. This shows, both a comparison of a larger size open review pattern against the traditional peer-review with few (usually three) reviewers and also reliability of decision making for each type of peer-review.

In summary, there are three main research questions as follows:

1- What is quality of knowledge in research publications and how we can best define it to meet the objectives of peer-reviewing?

2- How reviewers' preferences can be best reflected and analysed in their decision making for a manuscript based on different knowledge quality attributes?

3- To what extent can knowledge quality assessment from peer-reviewing may show variation and reliability between larger scale (open peer review) and smaller scale (traditional peer-review) decisions?

1.4 Organization of this Thesis

This thesis consists of seven chapter, Chapter one (the present chapter) discusses the background of the study, research motivation, problem statement and finally research objectives and questions covered throughout this study. Chapter two gives an in-depth insight about what has been briefly discussed in the study background of chapter one. It explores and defines concepts from different perspectives and schools of thought that are used throughout this study and experiments. Chapter three illustrates the research process and methodology corresponding to the research questions and objectives that are proposed in chapter one. It also explains the roadmap of the design of experiment and data collection for this study. Chapter 4 presents the descriptive statistical data and analysis of the correspondents' preferences, background and decision outcomes collected from the survey of this study. It then analyses the dataset obtained from the both online survey and focus group study for different decision patterns based on preferences, knowledge quality attributes and decisions made in the experiment through the development of decision trees for three different scenarios and conducting face to face focused group interviews. Chapter 5 covers analysis from two types of decision tree validation methods in order to examine the effect of size and diversity in peer-review decisions. This is done through clustering and random replications of decisions. Chapter 6 discusses, in detail, the decision patterns and the output of the validation methods obtained from the three scenarios in chapter 4 and chapter 5 and how each scenario performs and can be interpreted. Finally, chapter 7 concludes with the findings and significance of this research and discusses the limitations of this study and a roadmap to some possible future works that can be investigated.

Chapter 2 : LITERATURE REVIEW

In this chapter, we start exploring the fundamental concepts related to our first objective on defining the quality of knowledge. We divide this term into epistemological concepts and investigate each element in the context of scientific publications. We then cover the literature and practices of peer-reviewing in scientific publications and how investigated knowledge quality attributes and their potential biases may affect the in peer-review process.

2.1 Defining Knowledge

In this section, we try to look from the epistemological perspective on how knowledge has been defined and categorised from different schools of thought and elaborate more on those definitions that are used throughout this study.

2.1.1 Knowledge Definition, Types and Taxonomy

Over the years, there have been various definitions proposed by many scholars depending on their understanding of the word “Knowledge”. The very generic term can be found by looking up in the Oxford Dictionary (2015). The Oxford Dictionary defines “Knowledge” as “Facts, information, and skills acquired through experience or education; the theoretical or practical understanding of a subject”. However, every now and then, there are some other definitions of “knowledge” that appear in scholarly publications by some eminent scholars. Davenport & Prusak (1998) explained knowledge as “A fluid mix of framed experiences, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information”. They branded this definition as the working definition of knowledge. Liebowitz (1999) listed another series of definitions of knowledge from various scholars. While Beckman (1998) sees knowledge as reasoning about information and data in problem solving, Wiig (1994) sees truth, beliefs, judgments and expectations and know-how as the major elements in defining knowledge. Alavi & Leidner (2001) described these differences, that are rooted in different schools of thought perspectives, in five categories; on whether the knowledge is observed as 1- State of mind 2- Object 3- Process 4- Access to information or 5- Capacity. Table 2-1 summarises these viewpoints on knowledge within each perspective (Alavi & Leidner, 2001).

Perspective	Observation
State of Mind	Knowledge as state of knowing and understanding
Object	Knowledge as an object to store
Process	Knowledge as the process of applying expertise
Access to information	Knowledge as a condition of access to information
Capacity	Knowledge as the potential to influence action

Table 2-1 Knowledge Perspectives (adapted from Alavi & Leidner (2001))

There are also different forms of knowledge used in the general literature. There have been various taxonomies to categorize knowledge (e.g. tacit vs explicit, individual vs.

social, propositional vs experimental etc.) (Alavi & Leidner, 2001; Nonaka & Takeuchi, 1995; Nonaka, 1994). There are also different types of knowldge for some of these knowledge taxonomies.

Table 2-2 shows some of these knowledge types by intersecting some of the knowledge taxonomies (Lam, 2000). With such a great diversity of taxonomies and types, Lemos (2009) has tried to summarise knowledge into three main categories described as: 1) knowledge acquired through knowing someone or something, 2) knowledge about the know-how and skills for something and 3) knowledge acquired through evidence about the truthfulness of a fact. On the basis of the last part of this statement, “Knowledge of facts and true propositions” is described as Propositional Knowledge.

	Tacit	Explicit
Individual	Embodied Knowledge (Practical experience)	Embrained Knowledge (Theoretical knowledge)
Social	Embedded Knowledge (Routines, norms)	Encoded Knowledge (Written rules, Procedures)

Table 2-2. Types of Knowledge (adapted from Lam (2000))

Lemos (2009) illustrated that the most significant characteristic of propositional knowledge compared with the other two categories is its endeavour to find the truth through observation and evidence. That’s why propositional knowledge is often a common point of debate among philosophers and in science. Knowing a specific form of knowledge for science, there are still various definitions of knowledge. The major variation in the definitions of knowledge is based on the constructs that need to be presented or satisfied to recognize the output as knowledge (Audi, 2011). For example, there is an argument that justification is not a necessary element in defining knowledge, which can be acceptable in the case of knowing someone by acquaintance or having skills in something. Meanwhile, with the propositional knowledge as the source of recognition in mind, the widely known definition of knowledge as a "Justified True Belief" (JTB) fits the conclusions that are often required in scientific fields (Arkoudas & Bringsjord, 2009; Audi, 2011; Lemos, 2009). It was the seminal epistemological approach of Polanyi (1974) that combined many of these aspects in order to illustrate the elements of propositional knowledge. Polanyi (1967) illustrated propositional knowledge as scientific knowledge which is defined as "Justified True Belief". He also gave an in-depth epistemological explanation to support each element of belief, truth and justification.

2.1.2 Propositional (Scientific) Knowledge as Justified True Belief

The JTB definition is made of three significant constructs namely Justification, Beliefs and Truth. Each of these should be identified in such a way to support JTB in the form of propositional (scientific) knowledge.

A) *Justification:* There are two main forms of justification. Lyons (2009) described them as *Propositional Justification* that is "determined by evidence one possesses" and *Doxastic Justification* which is "matter of a belief being based on sufficient evidence". In this study, the term Justification is referred to as Propositional Justification, as the meaningful form of presentation in the JTB definition for propositional knowledge. The Doxastic justification is the other form of justification that is embedded in making assumptions and is reflected in belief (assumption that does not necessarily need to be proven to be true).

B) *Belief*: is the second component in the JTB model that needs to be properly defined to correspond to propositional knowledge. Lemos (2009) suggests that belief can be seen as the Propositional Attitude, which is a relation between a subject and a proposition (Arkoudas & Bringsjord, 2009; Dietrich & List, 2010; Lemos, 2009). However, this kind of simple definition cannot respond to the truth-seeking nature of the JTB theory. Beliefs can either be dispositional, which is just a simple thought without any further consideration (e.g. belief about feeling warmer with a thicker clothes in winter) or belief about an entertaining proposition which needs to be further investigated and studied (e.g. your new light down jacket keeps you warmer than your old heavy woolen coat). The beliefs that entertain one's mind for further investigation are known as *Occurrent Beliefs* (Lemos, 2009). Propositional Attitudes in the form of occurrent beliefs can present the truth-seeking nature of attitudes that shape the true beliefs in the JTB theory.

C) *Truth*: The last component in the JTB theory is Truth. This might sound to be least controversial to many epistemologists (Zalta, 2012), however, its definition varies significantly with the form of knowledge. The definition of truth as the objective of the JTB not only carries an important role in aggregating all the three component of the JTB, but is also highly interrelated with the definition of knowledge quality (which is explained later in this chapter) (Audi, 2011). There are mainly three major theories about truth. The first is the external presence of the truth, which makes it independent from any proposition attitude (e.g. the sun shines on a clear morning sky is a true fact).

This form of truth is known under correspondence theory and is valid without the presence of any individual's belief. Second, the truth can be defined under coherence theory. In this theory, truth is the consequence of different but relevant propositions that result in the same statement (Truth). The truth is only valid if and only if all those propositional attitudes believe in the same statement. Third, the truth can be described under the pragmatic theory. In pragmatic theory of truth, the truth stays valid as long as it is useful in practice (either individually or in group) (Audi, 2011; Lemos, 2009; Nonaka & Takeuchi, 1995). While the ultimate truth can be illustrated under correspondence theory, nevertheless, it is always beyond the reach of science as this is seen as a perpetual truth-seeking phenomenon. The correspondence theory of truth also requires a consensus among different but relevant propositions. This requirement makes it very challenging in scientific fields. Firstly, it is not easy to measure the relevance of propositions (e.g. observing the socio-economical phenomena from either a social or economical perspective) (Audi, 2011; Lemos, 2009). Secondly, scientific truth may vary significantly in different contexts (e.g. the laws of classical and quantum physics). Considering all these limitations in science for finding the truth, this study adopts the pragmatic theory of truth for JTB theory. In this form of JTB, every assessment for a true belief about a proposition is justified individually.

The final goal of such scientific knowledge illustration is to aggregate the output as a decision on the quality of research works based on independent propositional attitudes received from peer reviewers.

2.2 Defining Quality

In this section, we briefly explain the term "Quality" proposed by various scholars and will later explain in-depth how quality has been perceived in defining "knowledge quality". This section, benefiting from the vast references from the related literature in this topic, tries to contribute as much as possible in defining the "Knowledge Quality" term in scientific studies with only a vague definition so far, and few studies have concentrated on this issue.

2.2.1 What is Quality?

If you google "Define Quality", the result will appear as "the standard of something as measured against other things of a similar kind; the degree of excellence of something." ISO (International Standard Organization) defines quality as "The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs". Juran (1992), gave a shorter definition of quality by defining it as "Fitness for intended use". In fact, there is no universal definition of quality. Reeves & Bednar (1994), have traced the history of quality over time to find out the strength and weakness of each definition. They came to the conclusion that the definition of quality may vary in different circumstances based on the factors that need to be taken into consideration. Such conditional view about quality has led some scholars in proposing quality based on various dimensions. Garvin (1996) has defined 8 dimensions[1] of quality for tangible products while Evans & Lindsay (2005) have set another set of 8 dimensions[2] for services. Maxwell (1992) proposed a shorter set of 6 dimensions[3] that, in his view, could fit health services better. One of the very popular

[1] Performance, feature, reliability, conformance, durability, serviceability, aesthetics, perceived quality
[2] Time, timeliness, completeness, courtesy, consistency, accessibility and convenience, accuracy, and responsiveness.
[3] Accessibility, equity, appropriateness, effectiveness, efficiency, and Social acceptability

dimensional views of quality was also introduced by Parasuraman, (2002) known as RATER[4] which proposed only 5 dimensions. Thus, the multidimensional view of quality also varies in both the number of dimensions and the definition of each dimension. However, the multidimensional view has found its way in defining quality in data-driven domains, information technology and knowledge-based systems (Y. W. Lee, et al., 2002; Pipino et al., 2002; Rao & Osei-Bryson, 2007). In this study, we use and explore more on this multidimensional view of quality for knowledge.

2.2.2 Differences of Quality in Knowledge, Information and Data

While a plethora of books and journal papers developing theoretical frameworks to measure and improve the quality of data and information have contributed significantly to the development of various topics like data mining and information retrieval in recent years, there are limitations and drawbacks on how to extend these measurements for the assessment of quality of perceived knowledge within a context. One major drawback in extending the definition of data and information quality to cover knowledge quality is the indifference is dealing with the quality factors underpinning various knowledge definitions. Although there is no unilateral agreement on differentiating data, information and knowledge (He & Wei, 2009; Shin et al., 2001), there are nevertheless dominant approaches to positioning knowledge as a competitive resource by differentiating knowledge from data and information through contextualization for decision making (Mancilla-Amaya, Sanín, & Szczerbicki, 2012; Shin et al., 2001). Indeed, Bohn (1998) argues that the quality concept in knowledge should extend beyond merely the quality definitions in data and information in order to embrace the decision making process within a context.

[4] Reliability, Assurance, Tangibles, Empathy and Responsiveness (between perceived and expected service)

According to Dey (2001), Context "is any information that can be used to characterize the situation of an entity". This definition of context suggests that, once knowledge is contextualized in a decision making process, related quality factors in data and information are characterized by knowing about that context (Zimmermann, Lorenz, & Oppermann, 2007). While such definition of knowledge contextualization is definitely useful in differentiating the quality aspect of data and information from knowledge, additional research is needed to form a sustainable knowledge quality assessment in decision making.

2.3 Defining Quality for Propositional (Scientific) Knowledge

In this section, we illustrate the various aspects of quality in knowledge with consideration of the context from the multidimensional quality perspective. By integrating and summarising all the elements discussed in previous sections, this study proposes a knowledge quality assessment framework that will be used as a reference through the rest of this study.

2.3.1 Knowledge Quality

Knowledge quality is a key element for representing the value and sustainability of a knowledge-based process (Rao & Osei-Bryson, 2007). Over the past decade, there have been various attempts to resolve the issue of quality data which has ultimately lead to better quality information (Jarke et al., 1998). However, with no universal definition for quality, various definitions for quality have been proposed by researchers over the years based on the context of their studies (Mancilla-Amaya et al., 2012). As a result, researchers working on quality of knowledge have applied the same pathway in defining knowledge quality with the even more generic definitions for quality (Eppler, 2006; Huang et al., 1999; Lee et al. , 2007).

There have been extensive studies observing the quality of knowledge from different perspectives (Holsapple & Joshi, 2001; Owlia, 2010; Rao & Osei-Bryson, 2007). These studies were either conducted narrowly, based on quality attributes (Holsapple & Joshi, 2001), or widely explored through extensive reviews within multiple dimensions based on knowledge definitions (Rao & Osei-Bryson, 2007). In both cases, defining the quality of knowledge is associated with the proposed knowledge attributes observed in their own studies. While quality within a knowledge-based process can be traced in different facets of ontology (Burton-Jones, Storey, Sugumaran, & Ahluwalia, 2005; Jarke et al., 1999; Rao & Osei-Bryson, 2007), knowledge items (Aggestam, Backlund, & Persson, 2010; Holsapple & Joshi, 2001; Jarke et al., 1999; Rao & Osei-Bryson, 2007; Shin et al., 2001), knowledge retainers (Aggestam et al., 2010; Burton-Jones et al., 2005; Courtney, 2001; Holsapple & Joshi, 2001; Jarke et al., 1999; Rao & Osei-Bryson, 2007; Shin et al., 2001) and knowledge usage (Alavi & Leidner, 2001; Burton-Jones et al., 2005; Jarke et al., 1998; Kwan & Balasubramanian, 2003; Rao & Osei-Bryson, 2007), finding a proper quality assessment framework for user generated content has been always a challenging task (Lim, Vuong, Lauw, & Sun, 2006) because, as mentioned earlier, despite data and information, quality in knowledge is heavily dependent on the context. The presence of context may not only affect selection of the right quality criteria, but can also affect the importance of each of those quality criteria (e.g. weight of each quality dimension) in different contexts. Thus, a sustainable knowledge quality assessment framework cannot be carried out without the characterization of knowledge attributes that need to be measured. Such characterization usually takes place by providing contextual descriptions on those knowledge attributes that originated from some form of knowledge resources, as will be explained in the next section.

Maxim and Sluijs (2011) defined three major domains for knowledge quality assessment – namely the *substantive dimension*, *contextual dimension,* and the *procedural dimension.* With these dimensional definitions, their study on the quality of knowledge involved defining the parameters and causal relationships (Substantive), imposing constraints on knowledge generation through various resources (Contextual), and focusing on how and to whom the knowledge was generated for (procedural). Another study, done by Yoo, Vonderembse & Ragu-Nathan (2010) classified knowledge quality in terms of a) *intrinsic* quality where the knowledge inherently has quality, b) *Contextual* quality, which the knowledge can be reflected in the context, and c) *Actionable knowledge* quality which refers to the extent to which knowledge can be applied to different tasks (aka generalizability).

Table 2-3 outlines some proposed quality attributes associated with various attributes of knowledge quality (Rao & Osei-Bryson, 2007). As there are no de facto attributes for quality within knowledge-based processes, some significant ones (based on their appearance in the literature review) are represented in Table 2-3. By checking their illustration of measurement, it can also be seen that some of these attributes borrow their measurement approach from data and information concepts (which inherently affects the generation of a knowledge item (Holsapple & Joshi, 2001; Jarke et al., 1999).

Attribute	Measurement
Accuracy	The degree the knowledge can be verified to be true (verification)
Consistency	Number of contradiction occurrences among extracted items or with predefined rules
Currency	Time of verification
Interoperability	Non-interpretability rate of explored knowledge items within each knowledge discovery cycle.
Degree of context	The quantity of context available
Degree of details	Level of details provided (depth) plus number of different task usage (breadth)
Degree of Importance	Significance of the tasks the knowledge item is used for based on tasks' citation frequencies
Sharing	Number of times the different communities of practice (COP) have accessed the knowledge item
Usefulness	The amount of new knowledge generated based on the usage of a particular knowledge item
Volatility	The frequency that a certain knowledge item changes

Table 2-3 Knowledge item quality attributes and measurements (Adapted from Rao & Osei-Bryson, 2007)

Nonaka (1994) proposed a key viewpoint about the quality of knowledge. He divided the quality of knowledge for both tacit and explicit knowledge. For tacit knowledge, Nonaka showed that the quality of tacit knowledge originated from two factors; First, in the "variety" of an individual experience and second in the "knowledge of experience". (Nonaka, 1994). However, for explicit knowledge, he followed the Justified True Belief (JTB) definition of knowledge, which was discussed earlier in this chapter, as the propositional (scientific) knowledge that this study supports. Under the definition of JTB of explicit knowledge by Nonaka (1994), he put emphasis on the justification part as the determining element of Knowledge Quality. He believed that judging truthfulness (please refer to earlier part of this chapter for definition of justification and truth used in this study) should involve some standards (criteria). This illustration squarely fits the multi-dimensional quality perspective based on the knowledge context that has been chosen in this study. Garvin (1984) agreed that in the majority of studies in knowledge quality the multidimensional representation of quality is represented by assessing the quality from different perspectives.

As mentioned earlier in the previous section, with no consensus on the definition of quality for knowledge, the majority of the studies on quality have taken multi-dimensional perspectives. However, such dimensional observations still could not give a holistic view about selecting the most suitable dimensions (due to the difference of the context). Seawright & Young (1996) defined value-based quality as "excellence, or fitness for use, at an acceptable price". This was the extended definition to Juran (1992)'s quality definition of "Fitness for intended use" that provided a new insight on assessing the quality of knowledge within a context (fitness for use). By integrating the multidimensional perspective of quality with the value-based definition of quality (Seawright & Young, 1996) mentioned above and the observation of knowledge as a process (as one of the perspectives of defining mentioned earlier in this chapter), the quality of knowledge can be illustrated as the added value in the knowledge creation process. Moreover, "acceptable price", as mentioned by Seawright & Young (1996), can provide an insight on qualitative, quantitative or hybrid assessment approaches for the quality of knowledge which is further explained in the next section.

2.3.2 Knowledge Resources and Assessment of Knowledge Quality

Knowledge resources are the inputs for assessing quality of knowledge. According to Holsapple & Joshi (2001), knowledge resources can be identified in the form of either a *Schema*, which is dependent on some agreed templates as accepted norms (e.g. a scientific paper format in a discipline) or, in the form of *Content*, which is independently generated, disseminated and absorbed by the users. Rao & Osei-Bryson (2007) stated that the user (i.e. human factor) plays an important role in the formation and sustainability of knowledge quality assessment practice. While in the multidimensional assessment of data and information quality, applying a series of quantitative methods may give some valuable insights (Lee et al., 2002; Pierce, Kahn, & Melkas, 2006), these quantitative measures may only be helpful in assessing the quality of the schema (e.g. a scientific paper format). However, user generated content (as an independent knowledge resource) is a socially-constructed phenomena within a context (Popper, 1972). Thus, with regard to the need for contextual descriptions to support the assessment of knowledge attributes that originate from independent knowledge resources (Content), elimination of human assessment from the knowledge quality assessment may not only affects the knowledge quality, but also any changes in the context would render the existing knowledge detached from their original quality assessment attributes. (Huang et al., 1999).

In order to better clarify this concept, we give an example related to the topic of this study. There are many plagiarism checking software that may use many complicated data mining algorithms to find out any other similar work of a submitted manuscript, or various editorial systems that are designed to insure the submitted work complies with the required format. Nevertheless, these tools are only checking the Schema element of knowledge resources. The SCIgen project[5] of MIT is a famous show case of computer-generated content which perfectly fits the schema (e.g. conference paper template and format) (Labbé & Labbé, 2012; Zhuang, Elmacioglu, Lee, & Giles, 2007). The computer generated paper output of this project could get accepted for a conference as it perfectly fits the general requirements of a paper required in a conference. However, if there were any human expert reviewing the content, it could easily pick of all those missing points required in a quality paper.

Looking back to some attributes on the quality knowledge items mentioned in Table 2-3, for each of these attributes, the process of quality measurement can be either automated based on quantitative analysis derived from these measurements or qualitatively through human involvement in knowledge item reviews (Lim et al., 2006). The decision to choose the quantitative or qualitative approach or both (hybrid human-computer participation for quality item extraction) depends on many factors such as the nature of the content and data (Lim et al., 2006), locus of knowledge item (Holsapple & Joshi, 2001) or typology of context (Social vs. individual) (Memmi, 2008; Wenger, McDermot, & Snyder, 2002; Wenger, White, & Smith, 2009). Each of these evaluation methods may present a different level of accuracy as the nature of a knowledge item usually varies or evolves during the elicitation time. For example, in the mathematical and engineering context, a quantitative method may accept or reject the accuracy of a proposed method in a scientific work but needs a human expert to grade its novelty. For a business context, the feasibility maybe more important even if the proposed engineering method may sound novel and valid. Hence, these changes

[5] https://pdos.csail.mit.edu/archive/scigen/

of context sometimes make it even more difficult to warrant the measurement concepts applicable to every context.

2.4 Knowledge Quality in Peer-review Process

In previous sections, we discussed about the definitions of "Knowledge", "Quality" and "Quality of propositional (scientific) knowledge". By defining scientific knowledge as the Justified True Belief (JTB) and how quality of propositional knowledge has be observed from the multi-dimensional perspective by taking into consideration the context (fitness for intended use), we explore these aspects further in the domain of the peer-reviewing process for scientific papers. In this section, we focus on how quality of knowledge is assessed for publication of scientific works based on the concepts defined so far.

2.4.1 Introduction to Peer-review Assessment Process

Peer-review is the most popular method in today's scientific work assessment for publication, where a research document is assessed by number of scholars that independently evaluate the quality of the research work to decide whether it is suitable for publication or not (Bornmann, 2011) . In fact, peer-review, as an assessment method for the quality of scientific publications can be traced back to the 16th century (Biagioli, 1998; Spier, 2002). Over the years, not only has peer-review helped the scientific community to share values and norms, but also endorsed any scientific published work among the members of that discipline. Rowland (2002) summarized the four key objectives of peer review as: "1) Dissemination of current knowledge 2) Archiving the canonical knowledge base 3) Quality control of published information and finally 4) Assignment of credit and priority for their work to the authors." As it can be seen, quality control is one of the main objectives that needs to be achieved through the peer-review process. There are different types of peer-review (which will be explained in the next section) and each type may have a different process to follow based on the number of reviewers, with or without initial monitoring, revision policies etc. Figure 2-1 shows the peer review process adopted from Elsevier as one of the most reputable publishers for scientific publications. Such peer-review practice might vary among different publishers on the number of steps, the number of reviewers involved and actions taken on major and minor decisions, but according to Rowland (2002), the main body of the peer-review process with an editor and two referees (reviewers) has been left intact since the Second World War II.

Peer-review is a time-consuming and in most cases an unpaid task for reviewers; thus, it is basically seen as a voluntary work for those scholars who want to make a contribution to knowledge in their domain, or in some cases, receive recognition from peer scholars by being part of a reputable academic journal. The decision made after the peer review process for academic publications are usually Accept, Reject, Major Revision or Minor Revision. Based on these four common decisions types, Rowland (2002) mentioned that the amount of work needed in rejecting a paper may not be anything less than accepting it. In summary, with all the time and effort needed to have a scientific work peer-reviewed, it is still the most popular way of assessing knowledge quality for academic publications in the 21st century.

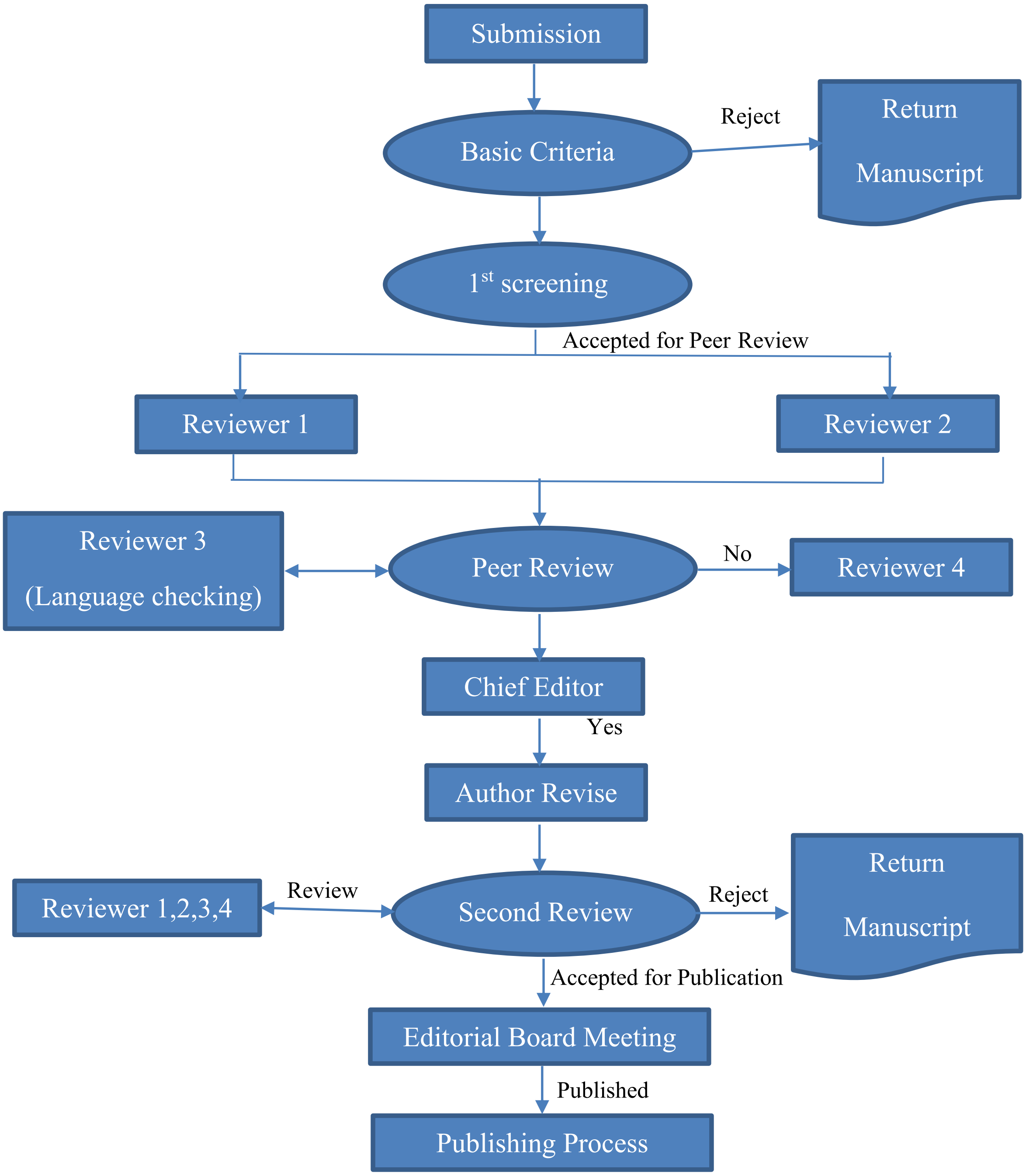


Figure 2-1 Peer-Review Process (adapted from Elsevier)

2.4.2 Types of Peer Review

As we discussed earlier, the peer-review process may vary in the number of reviewers and the number of steps. However, there are other major elements that not only affect the algorithmic flow chart of the peer-review, but also the elements involved in human decision making. In general, there are three major types of peer-review known as *Single Blind Review*, *Double Blind Review* and *Open Review* (McCormack, 2009).

1) Single Blind Review: In this peer-review type, the identity of the reviewers are hidden from the author(s) but reviewers are aware of the identity of the author(s). This may help reviewers to freely comment and criticize a research work without being influenced by possible contingent future reprisals from the author(s). In fact, this method is the most commonly used method for peer reviewing (Ware, 2008). Some scholars support this peer-reviewing method while some others disagree with this practice (Blank, 1991; McNutt, 1990; Suárez et al., 2012)
2) Double Blind Review: In this peer-review type, the identity of both reviewers and authors are hidden from each other. Some suggest that this method of peer-review creates less personal bias due to mutual anonymity (Budden et al., 2008; Jadad et al., 1996; McNutt, 1990; Nicholas et al., 2015), while other scholars believe the peer-review in its current format is flawed (Horrobin, 1990; Smith, 2006)
3) Open Review: In open review, the identity of both reviewer and author(s) are known to each other. With the emergence of the internet and later academic social media, the open review process is known to be the latest trend and has gained more attention in recent years (Ford, 2013; Hardaway & Scamell, 2012).

In the extensive literature review and report on peer-review from Ware (2008), he summarized the proportion that each of these peer-reviewing methods contribute to academic publications. Figure 2-2 shows a summary of the percentage for each of these peer-reviewing methods that has been used by editors or experienced by the authors. The post-publication review is in fact a kind of a commentary format that is not officially categorized under peer-review, but being used to encourage ideas that can be derived from the published work.

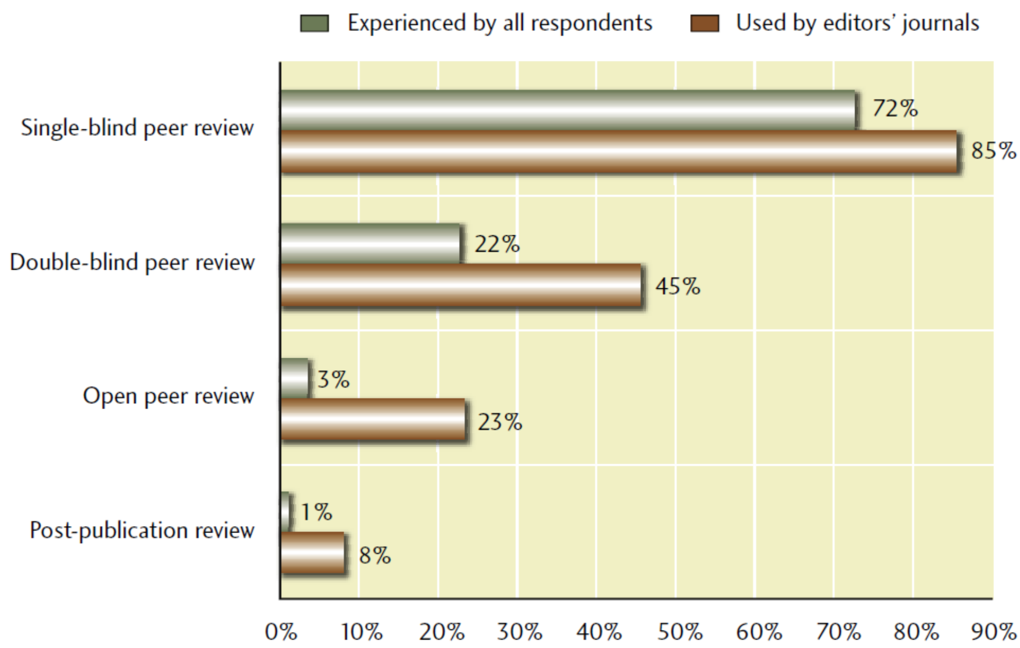


Figure 2-2. Types of Peer Review and Percentage of Usage (adopted from Ware, (2008))

On the other hand, Ware (2008) showed that a popular peer review is relative for authors and publishers. As Figure 2-3 shows, while single blind peer-review is the most commonly used method used in the scientific world in general (most popular when covering all scientific disciplines), the double blind review has received more support in terms of effectiveness.

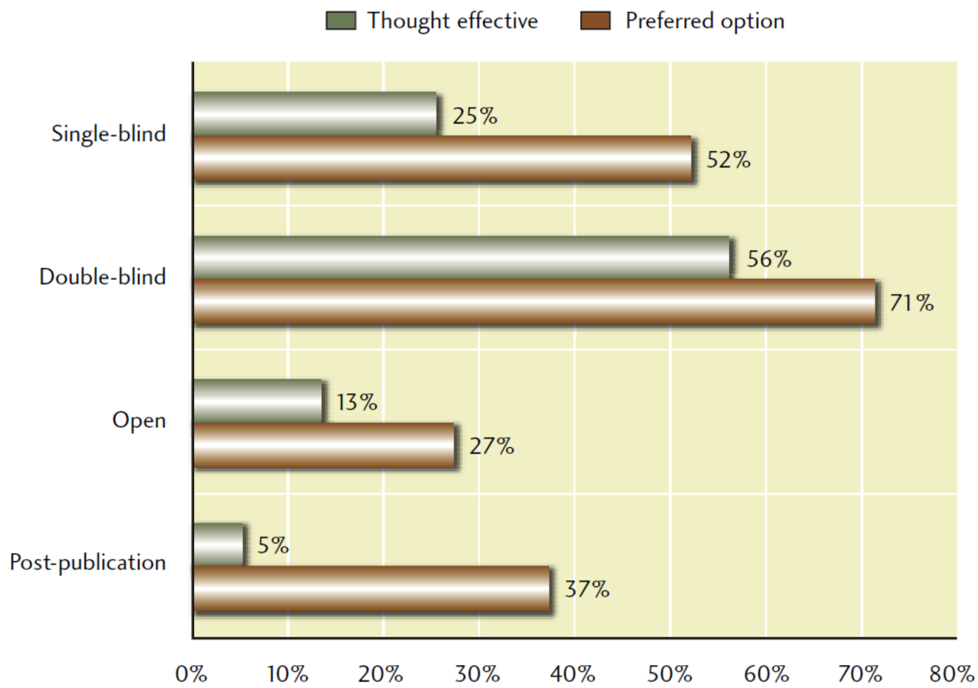


Figure 2-3. Effectiveness of Each Mode of Peer-Review (adapted from Ware, (2008))

While there is no unilateral agreement on which type of peer-review is the most suitable way of scrutinizing academic publications, there have indeed been some significant changes with the emergence of World Wide Web and social media. The scientific connectivity has taken the open review process to what is known as *Public Open Review* (Bornmann & Mungra, 2011). Ford (2013) described this form of Public Open Review as *Crowdsourced Review* in which a community of scholars from a discipline can give feedback for improvement on the proposed research. Such crowdsourced reviewing can also contribute to the advancement of knowledge and in strengthening the communities of practice in any scientific discipline (Ford, 2013). One of the challenges the public open review or so called crowdsourced review faces is the restricted accessibility to academic publications imposed by rights access from the publishers. Such a restriction has influenced the tendency to choose open reviews on a larger scale (Armbruster, 2008). Thus, with much research on the benefits and weaknesses of each of these peer-review methods, there has been little study to measure the scale of openness with many reviewers in comparison to those traditional single or double blind approaches. Ware (2011) suggested a series of challenges and opportunities that can be explored in opening up the peer-review process rather than strictly following traditional practices.

2.4.3 Biases in Peer-review

Irrespective of the type of peer review, the reviewer's judgment, as the core part of the reviewing process, is prone to biases due to the subjectivity of the decisions made by human beings at any level of expertise. There has been a lot of research on the origin of the biases that can influence decision making in the peer-reviewing process. In fact, the majority of articles criticizing peer-reviewing in its current model are focusing on the decision biases (Blackburn & Hakel, 2006; Langfeldt, 2006; C. J. Lee et al., 2013; Smith, 2006). In this part, we explore the origin and types of these biases.

In the peer-review process, as briefly discussed in the introduction chapter, there are two major elements that can create bias. The first one involves elements that are part of the literature review assessment and are based on the scientific measures that are used from the reviewer's perspective, and known as proxy measures (Goodman et al., 1994; Lee et al., 2013; Van Rooyen, Godlee, Evans, Smith, & Black, 1999). In this type of bias which is directly related to the true quality of assessment (Lee et al., 2013), the reviewer may select scientific measures and yardsticks to assess a piece of a scientific work which might not really be suitable measures for the research work to be tested against. Each reviewer, based on his or her expertise and level of knowledge in the scientific domain, may take into consideration a different series of measures for a submitted manuscript. Thus, the response from the reviewers may be significantly different depending on their expertise and the measures they have used for over a single manuscript. This is usually referred as low *inter-rated reliability* among reviewers (Bornmann & Daniel, 2010; Jackson, Srinivasan, Rea, Fletcher, & Kravitz, 2011; C. J. Lee et al., 2013). However, Schultz (2010) believed the empirical studies shows that the number of reviewers may not affect the final decision. Exploring further on this uncertainty is one of the objectives of this study.

The second type of bias originates from the character of the reviewer and factors outside the formal peer-reviewing process (Lee et al., 2013). This has also been one of the strong points of supporting the double blind review too. (Blank, 1991; Budden et al., 2008; Resnik & Elmore, 2015). However, Smith (2006) challenged the idea in that there is a great chance that the content of a manuscripts reveal or give clues to the full or partial identity of the author. In fact, with the power of search engines, it is not easy to assure the full anonymity of the authors of a submitted research work to the reviewers.

According to Lee et al. (2013), the bias that originates in the reviewer's characteristics can be categorized as either the bias that comes from disclosing the identity of the author to the reviewer or as a function of the reviewer's preferences for the content. Table 2-4 illustrates some of these extrinsic peer-review biases and also summarizes the different types of biases that be may derived from the above mentioned two categories.

Origin of the Bias	Bias Type	Description
Based on the author’s identity	Prestige Bias	The preferential evaluation based on the reputation of author in a scientific domain (Matthew effect) (Merton, 1968)
	Affiliation Bias	The bias that comes from formal or informal relationships between a reviewer and an author (Sandström & Hällsten, 2008)
	Nationality Bias	The bias that favors authors located in the same country or region (Link, 1998)
	Language Bias	The bias that favors English to non-English speaking countries (Ross et al., 2006)
	Gender Bias	The bias that favors one gender over the other (Budden et al., 2008)
Based on the content	Confirmation Bias	The bias that has “tendency to gather, interpret and remember evidence in a way that affirms rather than challenge one’s already held belief” (Nickerson, 1998)
	Conservatism	The bias against groundbreaking and innovative research (Braben, 2004)
	Bias against Interdisciplinary Research	The bias where the reviewers prefer mainstream research (Travis & Collins, 1991)
	Publication Bias	The bias that involves tendency for journals to publish research demonstrating positive rather than negative outcomes (Bardy, 1998)

Table 2-4 Peer-Review Bias Originated from Review’s Characteristics (adapted from Lee et al. (2013))

The research of Lee et al. (2013) has provided a lot of references from different scholars for and against each of the biases mentioned in Table 2-4. Langfeldt (2006) also discussed peer-review biases over social dynamics with almost similar biases that can be observed in the reviewer's judgments and decisions in different formats. The presence of any of these biases to any degree can ultimately affect the quality of the peer review and any decision derived from it.

2.4.4 Assessment of Quality of Knowledge in Peer Review Process

In the early part of this chapter, we discussed the definition of knowledge, quality and quality of knowledge. We explained the definition of Propositional (scientific) Knowledge as Justified True Belief (JTB), the multi-dimensional aspect of quality and the fitness for intended use at an acceptable price, and how these definitions are applicable to the quality of knowledge. Jefferson et al. (2002) and Weber (2002) did extensive studies on the assessment of knowledge quality with relevance to medical science. They compiled a series of assessment factors, indicators and rating methods used in many related previous studies. In fact, a simple search in the literature on the quality of knowledge in peer-review process shows that medical science has initiated and pioneered this area due to the criticality of quality of assessment in this domain (Bornmann & Daniel, 2010; Jadad et al., 1996; Van Rooyen et al., 1999). Cokol, Ozbay, & Rodriguez-Esteban (2008) showed how manuscript retraction has been on the rise over the past decades due to flawed manuscripts submitted in medical field. Figure 2-4 shows the sharp rise of retractions over the past three decades.

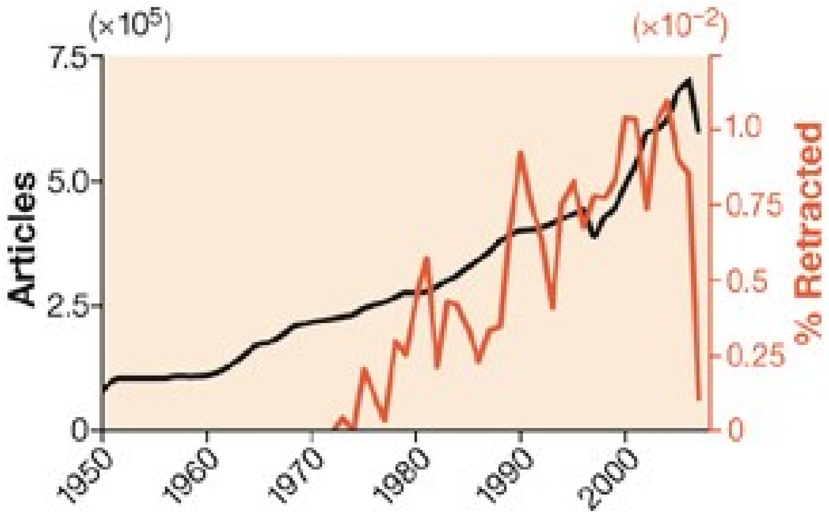


Figure 2-4. Articles retracted after submission in Medline (adopted from Cokol et al.,(2008))

In October 2013, the Economist magazine dedicated a special issue on "How Science Goes Wrong," (2013) and a special report on reliability of research entitled "Trouble at the lab," (2013) that showed the extent and importance of quality assessment of peer-reviewing in medical science. Jefferson et al. (2002) compiled a series of quality criteria and their respective indicators from various clinical studies. As depicted in Table 2-5, the multi-dimensional approach to quality is present in the knowledge quality assessment as an output of the peer-review process, similar to the knowledge quality for knowledge-based processes that were discussed earlier in this chapter (Rao & Osei-Bryson, 2007). What can be inferred in comparing the two (quality knowledge in knowledge-based processes and in peer-review) is the high level of similarity in the quality dimension with only different indicators. These quality indicators mat vary slightly or significantly depending on the level of contrast between the contexts in which the quality of knowledge is assessed.

Definition	Ideal Indicator	Surrogate Indicator
Importance Study findings and major impact	Changes in the research outcome status	1- Citation rates 2- Media coverage 3- Correspondence
Usefulness Study contributes significantly to the scientific debate or knowledge on a subject	1- Contributes significantly within a systematic review of the topic 2- Narrows confidence intervals around estimates of effect	1- Contributes to non-systematic reviews or guidelines 2- Citation rates 3- Correspondence
Relevance Topic is relevant to the journal's aims and readers	Topic is relevant and consistent with the aims and readership of the journal confirmed by survey	1- Citation rates 2- Correspondence 3- Internet hit rates
Methodologically Sound Methods used are able to answer the study question	Study findings are replicated several times across different settings	1- Closeness of fit between methods and "evidence-based" methodological checklist 2- Correspondence
Ethically Sound Unnecessary harm to humans or animals has been avoided and study has been carried out and reported honestly	1- No divergence between reality and the report 2- Rights of humans and animals safeguarded 3- Privacy and informed consent maintained throughout 4- Raw data match presented data 5- Number preference check is negative	1- Study received ethical clearance 2- No complaints from participants 3- No duplicate publication
Completeness All relevant information is presented	1- There is no selective presentation of data 2- All relevant references are cited	1- The text is complete 2- The publication is complete
Accuracy Presented information is a true reflection of what went on	1- Measurements truly reflect magnitude of findings 2- Raw data match presented data 3- References are accurate	1- The figures add up 2. Corrections

Table 2-5 Sample Indicators of Quality in Editorial-review of Clinical Studies

(adopted from Jefferson et al., (2002)) Goodman et al. (1994) and Ware (2008) also did surveys on a series of manuscripts assessed by various quality criteria in peer-review process. Goodman et al. (1994) showed that, the improvement in the quality of a manuscript after peer-reviewing is correlated with its initial quality score. In other words, the lower the quality of the paper at the initial stage (when submitting), the less it has benefitted from the peer-review assessment for revision and resubmission. Ware (2008) showed that while around 64% claimed to have benefited from peer-review in identifying their scientific or statistical errors, the majority (around 94%) were those individuals who benefited from minor corrections on presentation, language usage and references. Jefferson, Rudin, Brodney Folse, & Davidoff (2007) argued that while the peer-review may increase readability of a manuscript, it makes little contribution to the validity of the submitted research. This somehow suggests that, if the reviewers are not convinced with the scientific ideas proposed in a manuscript, there may be less chance for the work to be published after a major revision. Benda & Engels (2010; Jackson et al., 2011) proposed the role of *Predictive Validity* that also affects peer-review decisions in these cases. Predictive validity in peer-review aims to predict the validity of a scientific work, on the long term for the scientific community. Thus, scientific topics that attract more citations in a given period of time have a higher chance to be accepted. As discussed earlier in the previous section on peer-review biases, predictive validity can also be categorised as part of the reviewer's characteristics (extrinsic factors) which influence the peer-review process. In order to increase the accuracy of predictive validity, Benda & Engels (2010) suggested creating group reviewing with more reviewers in order to allow more diversity in reaching a better prediction.

2.5 Working Definition of Knowledge Quality in Peer Review Context

Based on the topics that have been discussed throughout this chapter, we now summarise them into a working definition that can be applicable to the research framework of this study which will be later on explained in chapter 3. Throughout this chapter, we explored the definition of knowledge in the scientific context and discussed different approaches in describing quality from different schools of thought. We later identified the multi-dimensional approach to quality as the dominant approach used in the peer-review process. Thus, on the basis of what has been explored in the proposed concepts, a working definition of knowledge quality can be outlined as follow:

"Knowledge quality in a peer review process refers to the multi-dimensional examination of the quality factors (attributes) that are considered to be the critical elements in assessment of scientific knowledge within a specific context. Such quality attributes can vary significantly in terms of quantity and importance among different scientific disciplines. Nevertheless, the quality attributes within a specific scientific domain may or may not be considered in every peer-reviewer's judgment in arriving at a decision as a result of biases that are originated from the peer-reviewer's personal preferences."

From the explanation in the above paragraph, it suggests that, the selection of these quality attributes can vary among different scientific disciplines and may not be generalizable from one scientific discipline to another. This requires a group of experts from each scientific discipline to identify those quality attributes that are considered to be significant within their domain of expertise (Ettenson & Shanteau, 1987).

2.6 Summary of Literature Review

In this chapter, we mainly discussed three major topics. We started this chapter with the epistemological views on the definition of knowledge, and continued to discuss about the propositional (scientific) knowledge and how it can be described as Justified True Belief (JTB), used in this study. We then discussed different propositions of quality and how the multi-dimensional view of quality can comply with its fitness for intended use in all contexts. In the third part of this chapter, based on the definition of scientific knowledge and quality, we proposed the quality of knowledge in scientific contexts. Finally, in the last part of this chapter, we investigated how quality of knowledge has been assessed in the peer-review process, so far, through an extensive literature review and identified the biases that can have an impact on the quality of knowledge assessment in the peer review process. We also mentioned how the multi-dimensional view of knowledge quality fits into the experiments and studies that were conducted on knowledge quality assessment in the peer-review process and finally provided a working definition of knowledge quality in the peer-review context.

In the next chapter, based on the definitions and concepts discussed and developed in this chapter, we proceed to propose a conceptual model and research framework to explore the behavioural patterns of reviewers and develop a methodology to test the preferences and biases that can affect the decision outcome in a peer review process. We also extend the methodology to compare two different modes of peer review (single blind vs. public open) in order to find differences and similarity in the relationship between the number of reviewers and decision outcome we also evaluate the predictive validity which has been discussed at the final part of this chapter.

Chapter 3 : RESEARCH METHODOLOGY

The methodology developed for this study is based on the multi-dimensional perspective of quality of knowledge discussed in the previous chapter. The methodology proposed here aims to find patterns of reviewer's preferences and characteristics that leads to various decisions in the peer review process based on the various knowledge quality attributes. Moreover, it will provide a decision comparison model to evaluate the impact of the number of reviewers (decision inter-subjectivity) on a final decision. The decision comparison will also be made between the quantitative approach obtained from the decision trees and the focus group face to face interviews where the decision trees are built manually by the participants. Two separate papers are used in this study for each mode of analysis (quantitative and qualitative face to face) in order to get a better understanding about the extent generalizability of the experiment.

3.1 Exploratory Research

The goal of exploratory research is to explicate fundamental concepts by asking individuals who are knowledgeable about a topic or process (Van Selm & Jankowski, 2006). With knowledge quality as a relatively vague concept (Melkas & Harmaakorpi, 2008; Pierce et al., 2006), many scholars have tried to determine the best quality attributes that fit the purpose of the context of their studies. Rao & Osei-Bryson (2007) proposed a series of quality dimensions by researching these attributes from the literature to identify different aspects of knowledge management systems. Weber et al. (2002) and Jadad et al. (1996) explored knowledge quality criteria that are taken into consideration in medical studies. Bornmann & Daniel (2010) investigated the scientific knowledge quality in chemistry and physics. In this study, we select another domain of scientific disciplines by choosing peer-reviewing for Knowledge Management (KM) as the context of the exploratory research in this study. Figure 3.1 shows the constructs of the research framework for this study.

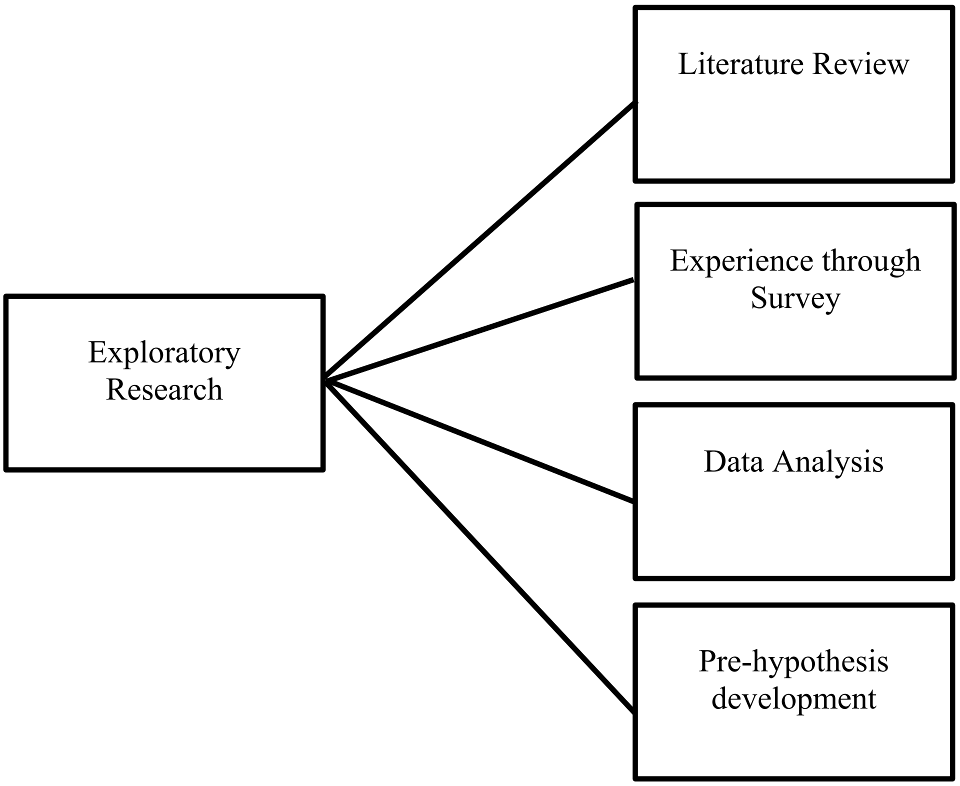


Figure 3-1 Research Framework

3.2 Descriptive Approach

When theories on a topic are not well-established, having some indicators gives a better understanding because these indicators can be tested, compared and also help in evolving theories (Black, 1999). In studies related to knowledge quality in the peer-review process, many scholars have set these indicators from the evaluation of a series of knowledge quality attributes (Bornmann & Daniel, 2010; Levis et al., 2015; Weber et al., 2002). Such indicators are part of the descriptive approach to find patterns in the data collected from knowledge quality attributes commonly used in peer-review that are perceived and assessed by the reviewers. The structured survey with close-ended questions is a suitable way for descriptive analysis of data as it provides comparable measures (Kothari, 2009).

There are several exploratory research studies with descriptive approaches to determine a reviewer's behaviour based on his or her preferences (Bornmann, 2008; Godlee, 2000; Goodman et al., 1994; Lamont & Huutoniemi, 2011; Van Rooyen et al., 1999). In this thesis, a structured closed-ended survey gives a quantified and comparable output for the data analysis inside and among each of the developed scenarios.

This thesis also explores the level of agreement among reviewers at different levels and examines the predictability of decisions based on a set of knowledge quality variables. Such agreement is related to the topic of inter-subjectivity in decisions. Smaling (1992) believes that the inter-subjectivity of attitudes can be interpreted and inferred through a descriptive approach to data collection analysis in order to develop theories and hypothesis.

3.3 Reviewer's Characteristics and Bias Dimension

As discussed in chapter 2, a reviewer's characteristics can create bias in the peer review process. Such biases mostly originate from two sources; one from the awareness about author's identity and the other from the content (Lee et al., 2013). Benos et al., (2007) also reiterated these two sources of bias in the peer review process plus the element of the so-called "conflict of interest" due to personal and academic competition. In consideration of all these origins of peer-review biases and the types of peer-review discussed in chapter two, the only source of bias that cannot be easily managed is the bias comes from the preferred content for the reviewer. As, for the identity of the author(s) or conflict of interest, irrespective of how well the identity of the author(s) is kept hidden, it is still more controllable in the experiment (through applying the same experiment settings for all participants) than the content which is in the core of the peer-review process. Thus, in developing our methodology for this study, we focus solely on the bias that originates from the content and keep the other elements to be consistent during the course of this study (similar to single blind and open review type). This is the bias that is usually categorized as ideological orientation and theoretical persuasion (Benos et al., 2007; Hojat, Gonnella, & Caelleigh, 2003). Figure 3-2 shows the focus of the methodology of this study from this bias dimension.

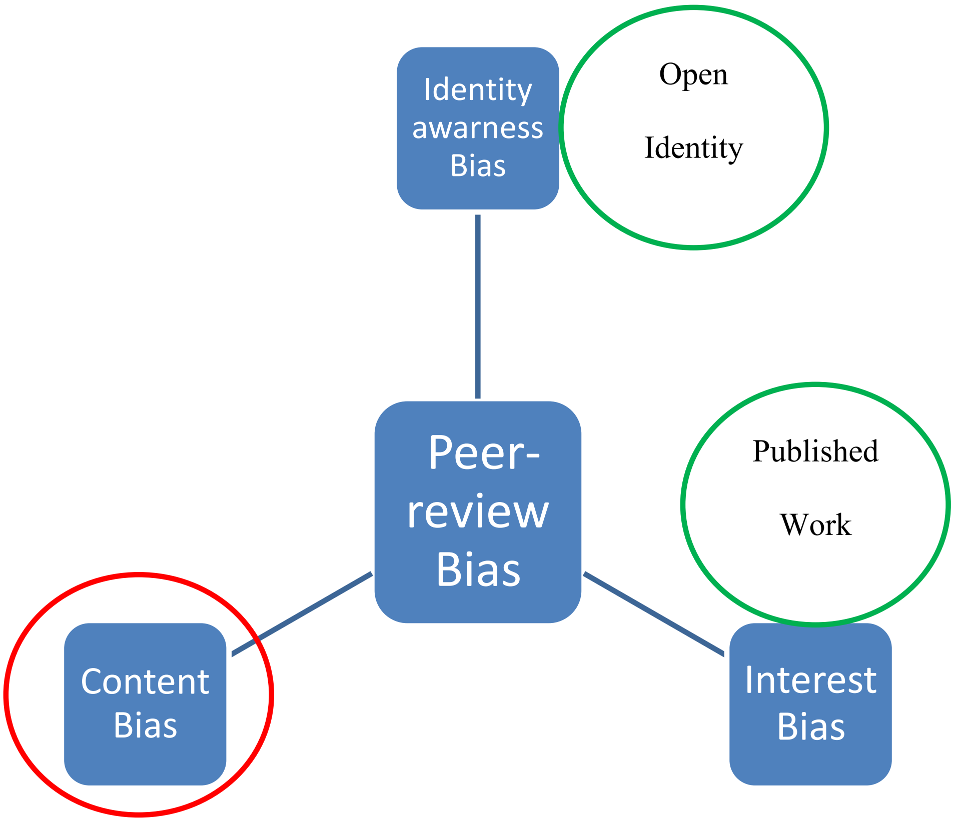


Figure 3-2 Bias Dimension of Methodology

Out of three main peer-review types (single, double and open), in two of them, the identity of the authors is known. In order to focus on the content bias in our methodology, we apply the same experiment setting by disclosing the identity of the authors to all participants. On the other hand, while bias from conflict of interest has not been the main concern in this study, nevertheless, this study chooses published work for assessment in order to ensure there will is no conflict of interest involved. Thus, the setup is similar to the post-publication review that was briefly discussed in the previous chapter. In this arrangement, what all the participants are needed to be rated is the content, and all the other biases (if any) remains similar to all the participants. The details are further explained in this chapter in the research process.

3.4 Knowledge Quality Assessment Dimension

The quality of knowledge in peer-review assessment is discussed in details in chapter 2 (Jefferson et al., 2007; Jefferson et al., 2002; Weber et al., 2002). From the research studies that investigated knowledge quality assessment, almost all of them have applied the multi-dimensional perspective of knowledge within a specific context (e.g. clinical or medical science). Weber et al. (2002) investigated many of the previous studies on the impact of review quality and acceptance on satisfaction which shows the multi-dimensional approach to quality assessment with different selection of quality criteria in different studies. In this study, instead of arbitrarily devising and selecting these quality assessment criteria, the aim is to select the criteria that have been used by a great number of reviewers within a disciplines. For this reason, this study focuses on a specific scientific publication domain and extracts similar quality criteria that are being asked from thousands of reviewers on a daily basis in assessing manuscripts of many journals related to the same domain. As the answers to some of these quality questions are open ended and a descriptive approach has been taken in this study, we adopt wide scale scoring (0 to 100) for better distribution in reflecting all the responses in order to map some of these qualitative responses into quantitative measures for the descriptive approach. The details of implementation are explained in the research design and process section.

3.5 Decision Scale (Inter-subjectivity) Dimension

In the past decade, by benefiting from the World Wide Web and social media, the readership or reviewing of a manuscript has expanded from a few persons to anyone who searches and accesses that manuscript online. As discussed in chapter 2, not only has more transparency emerged in recent years in the research community (Hardaway & Scamell, 2012; Suárez et al., 2012; Ware, 2011) but also it has created the opportunity for some research works to be assessed openly online by communities of experts rather than a of few individual peer-reviewers. Schultz (2010), as discussed in the previous chapter, undertook one the very few studies on the impact of scaling in the peer-reviewing process. The inter-rater reliability in peer review that was discussed earlier (Bornmann, 2008; Lee et al., 2013) has showed great diversity of opinion for a few reviewers while Schultz (2010) observed some kind of convergence with more reviewers. In this study, in order to better understand this scaling effect, we use a decision tree validity method to compare the decision responses from the all respondents with the chunks of smaller clusters of the same population of respondents, as explained in Chapter 5.

3.6 Conceptual Framework

So far, we have defined the three dimensions that will be covered in this thesis. As Bias Dimension and Knowledge Quality Assessment dimension are inter-related, the first part of the study explores the quality rating from the reviewers based on their background and preferences (Chapter 4). The second part of this study is on the impact of the scaling of the peer review on the decision outcome (Chapter 5). Figure 3-2 shows the conceptual framework that is adopted throughout this study.

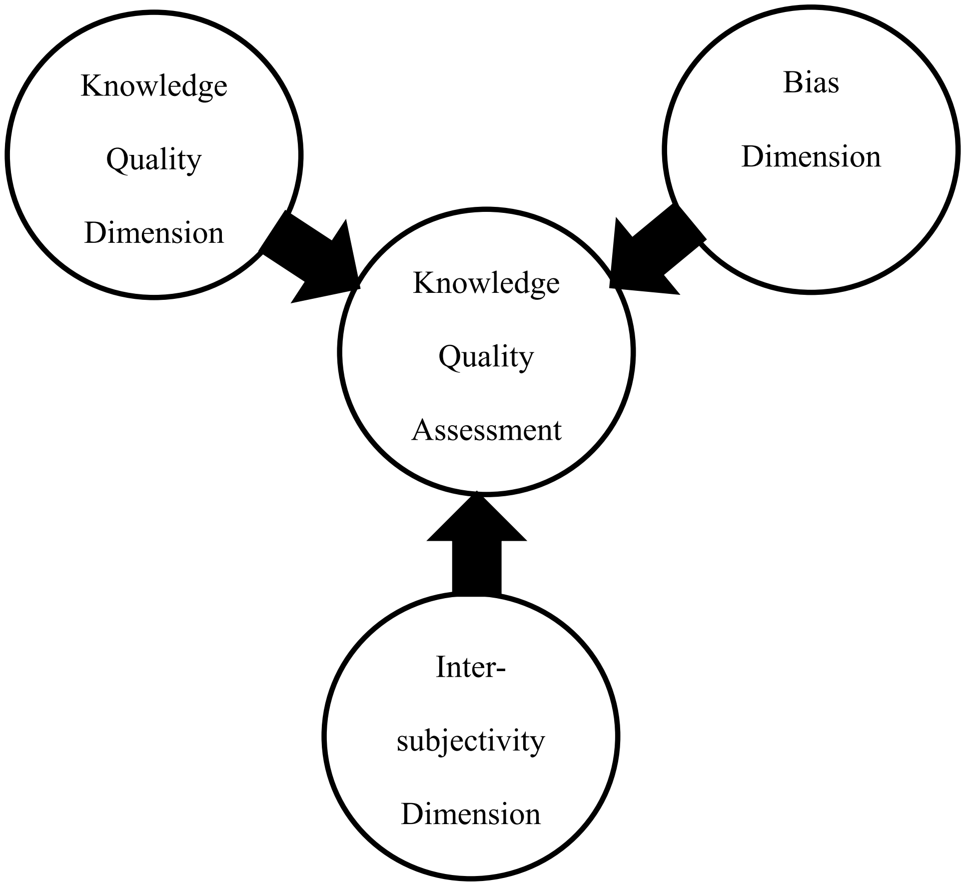


Figure 3-3 Conceptual Framework for Knowledge Quality Assessment

3.7 Research Design and Process

In this section we elaborate in detail all the steps in the conceptual framework in Figure 3-3. There are two main components in this study. The first one is the mechanism to collect reviewer's background inputs, preferences and biases and the knowledge quality rating, and the second element is to analyse the connection between different dimensions based on the reviewer's input and decisions. In order to achieve these conceptual framework requirements, this study has adopted a logic-based survey approach to collect the required data, both for bias and knowledge quality dimensions. Furthermore, the dataset collected through the survey is the input for a data mining software that will find the associations between decisions made by the reviewers and their quality rating inputs. Moreover, the data mining software evaluates the reliability of these decision patterns and test the reliability for different scales, which is the last dimension in our conceptual framework.

In order to have a better understanding about the generalizability and reliability of the decision tree structures using data mining in this study, a qualitative extension is also included using a focus group interview and manual decision tree building on the same set of quality attributes. However, a different source (paper) for judgment was used in order to eliminate dependency of the quality attributes from the same source (paper).

Figure 3-4 shows the details of each step and component in the research process.

3.8 Data Collection and Analysis

In order to collect data for analysis in all the three dimensions discussed in the conceptual framework, this study developed a fully flexible logic-based survey. A cloud-based survey tool named Qualtrics™, which is one of the most advanced logic-based online survey tools, was used as the survey tool. With consideration of the diversity of the reviewer's background and the time needed to read a manuscript for assessment, such a logic-based survey has significantly reduced the time needed to be spent in answering the survey questions.

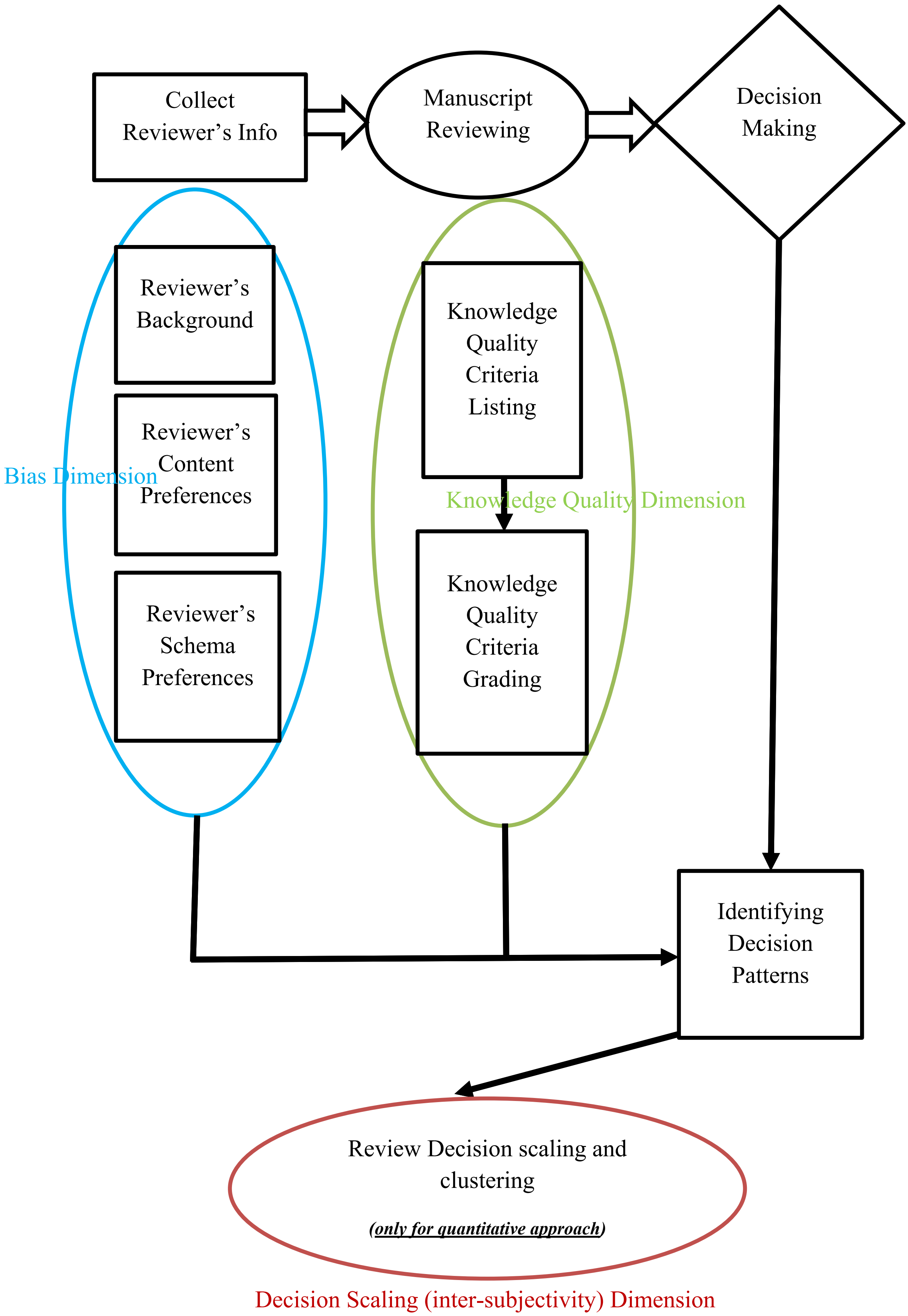


Figure 3-4 Research Design and Process

3.8.1 Survey Design for Quantitative Study

The survey entails three steps: A) Collecting Reviewer's background information and preferences B) Manuscript reviewing and quality rating C) Overall rating and decision making. The full survey questions and format are given in appendix B at the end of this thesis.

Collecting the reviewer's background was the introductory part of the survey, in which the respondents were asked about their academic position and whether they have had experience in peer-reviewing. In order to better focus on the scientific domain for assessment of knowledge quality in this study (similar to those experiments in clinical studies that were reviewed in the literature review chapter), the selected paper to be reviewed by the respondents was chosen from a published paper in the Knowledge Management discipline and from a related journal in the Knowledge Management domain[6]. The chosen paper was also in a more general topic that could be easily understood by the majority of the reviewers in this discipline. It should be noted that the main purpose of the selection of a paper for reviewing was to create a "point of reference" rather than content rating. In other words, as mentioned before, choosing a published work with known authors was aiming at controlling the bias dimension to focus on the reviewer's personal preferences and characteristics in assessing knowledge quality for a piece of research work.

In order to get better understanding about the background of participants for the selected discipline (Knowledge Management), we needed a list of reliable journals that usually receive most the submissions for the chosen domain. For this purpose, we adopted the list provided by Serenko & Bontis (2013) that provides details of 24

[6] The selected paper was an open access paper at http://dx.doi.org/10.1108/13673271311315196 chosen from the Journal of Knowledge Management

reputable journals in the Knowledge Management (KM) discipline, and provided the option for the participants to add up to 5 other journals for which they are reviewers or members of their editorial boards (further details are provided in Chapter 4).

The second part of the reviewer's information was about their schema and content preferences. As discussed earlier in Chapter 2, the schema refers to the preferences to any part of the format of the manuscript (e.g. more focus on the abstract) while the content preference is about what the reviewer sees as how a paper is delivering its message towards an objective. For the schema, this study follows the norms required by the journal of the discipline which are Abstract, Introduction and Literature Review, Research Framework and Methodology, Discussion and Conclusions and finally References. For the content preferences, this study borrows the criteria structure with some variation, that were suggested by the extensive studies in Gorman with journals and libraries (Gorman & Calvert, 2003; Gorman, 1999). These criteria are: 1- Advancement of Knowledge 2- Novelty in proposed ideas 3-Validity of the Proposed Methodology 4-Relaibility of the proposed hypothesis and research framework 5- Generalizability of the research framework and 6- Applicability of the research work. It should be noted that, in order to keep the progress of the survey more logical, the schema and content preferences questions were asked after the quality assessment phase of the survey in order to make it easier for the participants to understand the context of the proposed questions.

3.8.2 Survey Design for Qualitative Study

As in the qualitative approach, the survey was limited to a few number of participants who are known as experts in the domain, and the interview was conducted face to face, the reviewer background section was omitted in qualitative survey. (All participants in qualitative survey are reviewers of at least one reputable KM journal). In order to make a more reliable assessment with a better generalizability of the experiment, a different paper from a different journal was chosen from the list of journals that were mentioned in the quantitative approach (Serenko & Bontis, 2013). By offering a second point of reference, the qualitative study could provide a better understanding of decision pattern with less dependency in both content and schema(format) from a single reference point.

On the quality attribute part, similar to the survey for quantitative study, the same set of quality attributes for both content and Schema(format) was given to the focus group in the face to face interview, but instead of an online survey, the qualitative assessment of quality criteria was given to the participant on a paper after reading the selected paper in the KM domain.

In the qualitative approach, instead of using a software find the decision patterns based on the decision tree, we created a series of cards, each labelled with one of the quality attributes for both content and schema(format) that were also used in the quantitative approach. In this practise, we let the participants to build the decision trees manually based on the preferences and priorities. The priority of the preferences was built based on the vertical placement of the cards while preferences with similar priority were placed horizontally.

The survey questionnaire and the sample manual tree structure were shown in Appendix C at the end of thesis.

3.8.3 Knowledge Quality Assessment Design

After collecting the general information, the participants were asked to read the chosen sample paper (published and with known authors) and answer a series of quality assessment questions regarding the manuscripts. As there are various quality criteria that can be named and defined differently, this study selected exactly the same quality assessment criteria adopted by two major publishers in the discipline, namely Emerald and IGI-global. Some of these quality criteria from the selected sources are based on a Likert's scale and some are open answering. As this study follows the descriptive approach in identifying the decision patterns, a median approach was chosen. While the Likert scale is too narrow and open answer is too wide in terms of inputs, this study adopted a scale of 0 to 100 as a grading system. This enables the decision behaviour to be better classified on a wider range of answers. After grading each quality criteria, the reviewer gives an overall grade to the manuscript and chooses the final decision (reject, accept, minor revision or major revision).

3.8.4 Analysing Decision Patterns

With completion of the survey, the participant has assessed the background, schema and content preferences and a sample review of the paper with quality attributes that resulted in a decision. Such data is repeated for the other participants and provides a dataset with a series of variables as input in order to be mined for pattern discovery. This study uses the clustering approach of Decision Tree in SPSS™. With a decision being a categorical variable that falls into four categories (accept, reject, minor revision and major revision), Decision Trees are a popular technique in segmentation and classification in order to explain decision patterns by branching in a simple way (Kotu & Deshpande, 2015). The decision tree method also provides flexibility on a number of branches based on the number of variables (quality criteria in this study) and the sample size. Moreover, the decision tree excludes and compares decision cases from the same sample size which is the objective of this study in comparing the impact of scaling (number of reviewers). This can give an insight as to whether the decision pattern changes by reducing the number of reviewer or if the decision patterns stays the same, or to what extent is the output reliable.

In the next chapter (Chapter 4), we discuss all the survey outputs in detail for the data collected from the participants. We further explore the different associations that can be made between the first two dimensions (Bias and Knowledge Quality) with the decision variables to map out some of the emerging decision patterns. Chapter 5 explores the decision scaling impact through cross-validation and split-sample validation by dataset clustering and explores the decision reliability and changes with a different number of reviewers.

Chapter 4 : KNOWLEDGE QUALITY APPROACH TO DECISION ANALYSIS

In this chapter, we explain in detail the experiment design and implementation of the research process briefly discussed in Chapter 3. This chapter has three major parts with the first two parts related to the quantitative study and the third part on qualitative study that supports the findings of quantitative section. The first part explains the quantitative survey step by step, the data collected from the study, the background of participants and all relevant statistical reports in order to give a big picture of the participants' background, the context and preferences of the participants, their quality grading summary and their decisions. The second part of this chapter uses all the collected data obtained in the first part of the chapter to associate decisions with preferences and quality ratings. The third part of this chapter discusses using the same quality attributes to assess peer decisions among a group of experts in a manual way. For building up the decision trees, despite the second part of this chapter which discusses the use of analytical computer software, each individual create the tree manually based on their preferences (a sample in shown in appendix C at the end of this thesis)

4.1 Study on the Participant's Background in Quantitative Survey

The quantitative survey designed for this study aimed to target a specific group of people. The survey was designed in three sections. The first part was on collecting the general background of the participants, their peer review experience and the journal for which they normally do peer reviews. As the context of this study was chosen in the Knowledge Management (KM) domain, we tried to reach people who are either expert, professional or at least have some basic knowledge about the KM discipline. In the second part of the survey, the participants review a selected paper. As mentioned in chapter 3, this was a published paper with the authors' identity known that was published in a reputable journal of the discipline. This was to ensure that the experiment only focused on the bias originated from the preferred content where the participants review questions only focus on the content. This approach also matches most recent trends on transparency and open review practices that are discussed in chapter 2. After reviewing the manuscript, the participant should grade (from 0 to 100) the manuscript based on the quality attributes adopted from two major publishers in the KM discipline, Emerald and IGI-global. They finally should provide on overall grading and final decision for the manuscript.

The third and final part of the survey asks about both the schema (format) and content preferences of the participants. While the participant should grade each part of the schema (format) of the manuscript (abstract, literature review, methodology etc.), they should also rank the importance (preference) of each of these section. They should also define their preferences on what they usually expect to see in a paper in making a decision.

In total, 27 participants fully completed the survey. It should be noted that, with the amount of time needed to answer this survey, the participants' thoughtful answers should be higher than ordinary surveys in which they only respond to some questions without any pre-survey reading requirement.

The Qualtrics online survey with a logic-based design and analytics tools was used to conduct this study. The background info is explained in the following subsections.

4.1.1 Academic Background of the Participants

In order to maintain the quality of the respondents for this study, the survey was designed to be by invitation only. From more than 200 emails that were sent out, 27 people fully completed all the steps of the survey. Figure 4-1 shows the background of the participants in this study.

Category	Answer		Response	%
1	Professor		7	26%
2	Associate Professor		4	15%
3	Assistant Professor		3	11%
4	Senior Lecturer		1	4%
5	Lecturer		5	19%
6	Teaching Fellow		1	4%
7	Others (Please specify)		6	22%
	Total		27	100%

Research Associate
Project Associate
Researcher
Former lecturer

Table 4-1 Academic Background of the Participants

More than half of the participants in this study were Assistant Professor and above and all the participants had, at minimum, a post-graduate degree.

One of the other elements that is usually taken into consideration in the background study of reviewers is their geographic location (Gorman, 1999) as, usually, the considerations are format, content and other elements that may vary based on the standards that they have developed or adopted in their region. Figure 4-2 shows the geographic information of the participants of this study. This information collected was based on their current working place and not their place of birth. While the participants come from almost all regions, the majority are working in Asian and European academic institutions.

Answer		Response	%
Asia		14	52%
Africa		1	4%
Australia and Oceania		3	11%
Europe		7	26%
North America		1	4%
Latin America and Caribbean		1	4%
Total		27	100%

Table 4-2 Geographic Information of the Participants

4.1.2 Peer Review Experience and Expertise of the Participants

As Table 4-3 shows, 25 participants were actively involved in peer reviewing, while more than half of them were part of the editorial member of a journal. The peer reviewing experience of the participants is also shown in Table 4-4.

Peer-reviewer		Response	%
Yes		25	93%
No		2	7%
Editorial Member		**Response**	**%**
Yes		14	56%
No		11	44%
Total		25	100%

Table 4-3 Peer-reviewing Status of the Participants

Answer		Response	%
Less than 3 years		9	36%
3 to 5 years		5	20%
5 to 10 years		3	12%
More than 10 years		8	32%
Total		25	100%

Table 4-4 Peer-review Experience of the Participants

As mentioned in chapter 3, in order to provide a better understanding of the reviewing background of the participants in this study, we have adopted a series of popular journals in KM discipline from the study of Serenko & Bontis (2013) (Appendix 1). Table 4-5 shows this list of ranked journals and the number of participants who are either a regular reviewer or an editorial member of each of these journals. As it can been seen, the participants are either a reviewer or an editorial member of the KM journals, with 4 participants associated with A+ ranking journals, 7 participants associated with A journals, 16 participants associated with B ranking journals and finally 4 participants associated with C ranking journals. In addition to the journals listed in Table 4-5, Serenko & Bontis (2013) also listed a series of journals in which their core area might not be KM and therefore not ranked in the KM domain, but they are considered to be related journals to the KM domain. In addition, in the survey the participants were asked to provide the names of other related journals if those journals are not listed in the first table (Table 4-5). For both user-input and those related but non-ranked KM journals with at least one member from the participants in the survey, are listed in Table 4-6.

#	Journal	No. of Reviewers or Editorial Members	%	Rank
1	actKM: Online Journal of Knowledge Management	2	13%	C
2	International Journal of Knowledge and Systems Science	2	13%	C
3	Knowledge and Process Management: The Journal of Corporate Transformation	0	0%	A
4	Electronic Journal of Knowledge Management	1	6%	B
5	Intangible Capital	0	0%	C
6	Interdisciplinary Journal of Information, Knowledge and Management	1	6%	B
7	International Journal of Knowledge and Learning	1	6%	B
8	International Journal of Knowledge Management	2	13%	A
9	International Journal of Knowledge Management Studies	0	0%	B
10	International Journal of Knowledge Society Research	0	0%	C
11	International Journal of Knowledge, Culture and Change Management	0	0%	B
12	International Journal of Knowledge-Based Development	1	6%	B
13	International Journal of Knowledge-Based Organizations	0	0%	B
14	International Journal of Learning and Intellectual Capital	2	13%	B
15	Journal of Information and Knowledge Management	2	13%	B
16	Journal of Intellectual Capital	2	13%	A+
17	Journal of Knowledge Management	2	13%	A+
18	Journal of Knowledge Management Practice	0	0%	B
19	VINE: The Journal of Information and Knowledge Management Systems	7	44%	B
20	Knowledge Management & E-Learning: An International Journal	1	6%	B
21	Knowledge Management for Development Journal	0	0%	B
22	Knowledge Management Research & Practice	2	13%	A
23	The Learning Organization	3	19%	A
24	Open Journal of Knowledge Management	0	0%	C

Table 4-5 Participants Associated with Ranked KM Journals

#	Journal	No. of Reviewers or Editorial Members
1	Expert Systems: The Journal of Knowledge Engineering	1
2	International Journal of Technology, Knowledge and Society	1
3	Journal of Knowledge-Based Innovation in China	1
4	Knowledge and Information Systems: An International Journal	1
5	Knowledge and Innovation: Journal of the KMCI	1
6	Management Learning: The Journal for Managerial and Organizational Learning	1
7	Journal of Industrial Engineering	1
8	Industrial Management & Data Systems (IMDS)	1
9	The Asia-Pacific Education Researcher	1
10	Academy of Management Review	1
11	International Journal of Innovation and Learning	1

Table 4-6 Participants associated with non-listed and non-ranked KM Journals

In addition to the journals listed in Table 4-5, the participants were asked about their role in the KM related journals. One participant had the role of chief editor, 7 participants had the role of associate editors and 11 participants were part of the editorial member of those ranked journals.

In terms of expertise, Table 4-7 shows the claimed expertise of the participants. As can be seen, majority of the participants do have some level of expertise in KM or KM related domains.

Expertise	Response	%
Knowledge Management	18	72%
Artificial Intelligence	5	20%
Information Technology	4	16%
Data Mining/ Predictive Analytics	5	20%
Semantic Web and Technologies	4	16%
Computer Science/Software Engineering	4	16%
Business and Management	11	44%
Intellectual Capital	7	28%
Accounting and Finance	2	8%
Economics	2	8%
Intellectual Property	1	4%
System Thinking and Modelling	5	20%
Others	9	36%

Table 4-7 Areas of Expertise of the Participants

Up to this stage, the background information of the participants which was the first part of the survey has been elaborated. In the next section, we explore the second part of the survey which is on the participants' preferences, from both content and format (schema) of the manuscript, that they take into consideration while reviewing scientific research.

4.2 Study on Content and Schema Preferences of the Participants in the Quantitative Survey

In chapter 2, we discussed the biases that may affect the decision of the reviewer based on what he/she may expect from the content or any part of the schema (format) that may impact the decision of reviewer. In the survey, we asked the participant to rank their priorities for each of these preferences. For the content, as mentioned in chapter 3, we adopted (with some variation) the most important priorities collected through various studies by Gorman & Calvert (2003) and Gorman (1999). These priorities are: 1- Advancement of knowledge 2- Novelty in proposed ideas 3- Validity of the proposed methodology 4- Reliability of the proposed hypothesis and research framework, 5-Generalizability of the experiment and 6- Applicability of the research topic. By ranking between 1 to 6, Table 4-8 shows how many of the participants ranked each item at which priority level.

#	Preference	1	2	3	4	5	6	Total Responses
1	Advancement of knowledge	7	9	5	0	2	4	27
2	Novelty in the proposed ideas	16	4	2	2	2	1	27
3	Validity of the proposed methodology	1	6	9	9	2	0	27
4	Reliability of the proposed hypothesis and research framework	1	4	6	9	3	4	27
5	Generalizability of the experiment	1	2	1	4	10	9	27
6	Applicability of the research topic	1	2	4	3	8	9	27
	Total	**27**	**27**	**27**	**27**	**27**	**27**	**-**

Table 4-8 Preference of Content Objectives of Participants

As it can be seen, majority of the participants have selected the novelty of proposed ideas as their first priority in reviewing the content of a manuscript, and advancement of knowledge as the second priority. Table 4-9 shows the statistics of such content preferences in more detail.

Statistic	Min Value	Max Value	Mean	Variance	Standard Deviation	Total Reponses
Advancement of Knowledge	1	6	2.74	3.05	1.75	27
Novelty in the proposed idea	1	6	2	2.31	1.52	27
Validity in the proposed methodology	1	6	3.19	1	1	27
Reliability of the proposed hypothesis and research framework	1	6	3.78	1.87	1.37	27
Generalizability of the experiment	1	6	4.74	1.89	1.38	27
Applicability of the research topic	1	6	4.56	2.18	1.48	27

Table 4-9 Statistics for Content Preferences of Participants

As Table 4-9 shows, all of these preferences range between 1 to 6, meaning at least for one participant, each of these preferences has been chosen as the highest and lowest priority. Meanwhile, with the means closer to 1 and closer to 6, it can be deducted that novelty in the proposed idea and advancement of knowledge were chosen most frequently as the top priorities, while applicability of the research topic and generalizability of the experiment had a lower priority on average for the participants. From the standard deviation, it can also be concluded that validity in the proposed methodology is a more stable option with the lowest standard deviation as neither a high nor a low priority.

The participants were also requested in the survey to choose their preferences based on the schema (format) preferences, and their priorities in the sections of the manuscript that were considered to be the most crucial parts in making their decisions. The participants were given the most common format adopted by the majority of the papers published in KM discipline and in the two chosen publishers (Emerald and IGI). These items are: 1- Abstract 2- Introduction and Literature Review 3- Research Framework and Methodology 4-Discussion and Conclusion and 5- References. Table 4-10 shows how many participants rank their preferences between 1 to 5 on each specific section of a paper as their priorities in making a decision on a manuscript.

Preference	1	2	3	4	5	Total Responses
Abstract	5	3	5	8	6	27
Introduction and Literature Review	6	9	9	3	0	27
Research Framework and Methodology	13	5	3	4	2	27
Discussion and Conclusion	3	8	9	6	1	27
References	0	2	1	6	18	27
Total	**27**	**27**	**27**	**27**	**27**	**-**

Table 4-10 Preference of Schema (format) for Participants

As it can be seen, majority of the participants consider the research framework and methodology as the most important part of the paper while they see the references as the lowest priority in making their decisions. Table 4-11 shows the statistics of the schema preference by the participants in more detail.

Statistic	Min Value	Max Value	Mean	Variance	Standard Deviation	Total Reponses
Abstract	1	5	3.26	2.05	1.43	27
Introduction and Literature Review	1	4	2.33	0.92	0.96	27
Research Framework and Methodology	1	5	2.15	1.9	1.38	27
Discussion and Conclusion	1	5	2.78	1.1	1.05	27
References	1	5	4.48	0.8	0.89	27

Table 4-11 Statistics of Schema (Format) Preferences of the Participants

From table 4-11, it can be observed that all of the items have been chosen as lowest and highest priorities, except that the introduction and literature review sections have never been chosen as the lowest priority level. While references have been selected as one of the lowest priorities almost consistently, the introduction and literature review sections are seen as neither a very high not a low priority for the majority of the participants.

In the next section, we continue with how the quality assessment criteria were set and implemented in the survey and provide the statistics for quality assessment attributes similar to those discussed for both the content and schema preferences.

4.3 Study on Peer Review Quality Assessment and Decisions Patterns in Quantitative Survey

In this section, the survey design for knowledge quality assessment of the chosen paper is given first, and the summary of the responses based on the quality assessment questions is discussed in the second part of this section.

4.3.1 Peer-Review Quality Assessment Implementation

In chapter 3, we mentioned that in the multi-dimensional perspective of quality, quality criteria are very diverse for different studies and in different contexts. Weber et al. (2002) summarized a dozen studies in the same domain (clinical studies) and how these criteria differed, either by type or by definition, in different studies. In our study, in order to choose a more consistent approach in selecting knowledge quality assessment criteria, we have chosen the criteria from two major publishers, namely Emerald and IGI-global. As these quality criteria are used by thousands of reviewers every day to assess the manuscripts of most journals, it can give a better picture on how these quality criteria may be used in reaching a decision.

While some of these criteria are open and descriptive, and some are based on the 5-point Likert scale, in order to keep grading quantifiable, for this study while keeping a wider range of choices for the participants of the survey, we have made the grading as a bar between 0 to 100, similar to the grading of normal manuscripts. Figure 4-1 shows the sample quality grading bar for one for one of the quality criteria (Originality) in the survey. The participants can choose their desired grade by dragging and dropping the bar for each of the quality factors listed in the survey, after reading the manuscript. Finally, they can give an overall grade to the manuscript in a similar way and choose the final decision between the four options of Accept, Reject, Minor Revision and Major Revision.

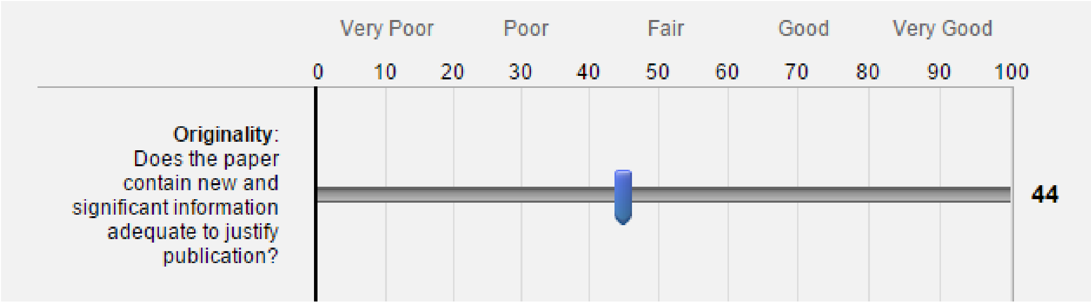


Figure 4-1 Quality Assessment Grade Bar

The quality criteria for this study was directly adopted from the online forms that were used by Emerald and IGI-Global at the time of this study. Table 4-12 and Table 4-13 show these quality assessment criteria adopted from these two major publishers for KM related publications. The criteria adopted by Emerald is more of a descriptive type that needs some participant input for open-ended questions in some cases, while the IGI-Global has a short multiple choice format, by merely mentioning the quality attribute and asking the reviewer to rate it on the 5 point Likert scale. In this study, both criteria from Emerald and IGI-global follow the same grading method of the 0-100 bar shown in Figure 4-1.

After responding to each of the quality criteria listed in Tables 4-12 and 4-13 (totally 15 questions) which are related to assessment of the content, there are also assessment questions on the schema (format) of the selected manuscript, similar to those listed in Table 4-11, with a slight expansion on the reference section.

Content Quality Attribute	Definition
Originality	Does the paper contain new and significant information adequate to justify publication?
Relationship to Literature	Does the paper demonstrate an adequate understanding of the relevant literature in the field and cite an appropriate range of literature sources? Is any significant work ignored?
Methodology	Is the paper's argument built on an appropriate base of theory, concepts, or other ideas? Has the research or equivalent intellectual work on which the paper is based been well designed? Are the methods employed appropriate?
Result	Are results presented clearly and analysed appropriately? Do the conclusions adequately tie together the other elements of the paper?
Implications for research, practice and/or society	Does the paper identify clearly between any implications for research, practice and/or society? Does the paper bridge the gap between theory and practice? How can the research be used in practice (economic and commercial impact), in teaching, to influence public policy, in research (contributing to the body of knowledge)? What is the impact upon society (influencing public attitudes, affecting quality of life)? Are these implications consistent with the findings and conclusions of the paper?
Quality of Communication & Language	Does the paper clearly express its case, measured against the technical language of the field and the expected knowledge of the journal's readership? Has attention been paid to the clarity of expression and readability, such as sentence structure, jargon use, acronyms, etc.

Table 4-12 Quality Assessment Attributes Used for Content in Emerald™

Content Quality Attributes		
Popularity of the Subject	Appropriateness for the journal	Adequacy of literature review
Quality of research design	Adequacy of data analysis	Contribution to the literature
Legitimacy of the conclusions	Practical/managerial significance	Clarity of presentation

Table 4-13 Quality Assessment Attributes for Content Used in IGI-global

Other than the quality assessment criteria adopted and listed in Table 4-12 and Table 4-13, a series of questions, shown in Table 4-14, were also asked about the schema (format) assessment of the selected paper for review in our study.

Schema Quality Attributes	
Abstract	Conclusions and Discussion
Introduction	Adequate References
Methodology	Accurate References
Results	Up-to-date References

Table 4-14 Quality Assessment Attributes Used for Schema (format) Screening

Finally, there was also an overall grade (between 0-100) for the whole manuscript, followed by the final decision as the last input from the participants. (Appendix B shows the survey with all the questions in each section).

4.3.2 Peer-review Quality Assessment Statistics

The selected paper for this study was chosen from an open access paper[7] repository of the Journal of Knowledge Management which is ranked as A+ journal in the KM disciplines (Serenko & Bontis, 2013). With the paper published and with the known authors' identities, the aim of this study was to focus only on the reviewer's bias toward the content while adopting a transparent peer-review process. Thus, as discussed in the preference questions earlier in this chapter, all the questions asked during the survey process were basically related to the understanding of the author in regard to the content in which its authors are known and the paper has been already published (similar to post-review practice). The 14 questions that needed to be graded on quality assessment were all those items that has been adopted from Emerald and IGI-Global listed in Table 4-12 and Table 4-13. Table 4-15 shows the summary of the statistics that were collected from the participants after reading the selected paper.

As seen in Table 4-15, there is great diversity in the opinions of the 27 participants in this study. Almost all quality attributes for content range widely between 0 to 100, meaning there is at least one participant for each of the quality attributes mentioned in Table 4-12 and Table 4-13 that are graded close to either 0 or 100. Looking at the standard deviation for all of these quality attributes, it can be seen that the variation in the diversity of opinions is relatively similar for all of the quality attributes for content, within the range of 0 to 100; with the lowest standard deviation for "relationship to literature review" and highest standard deviation for "clarity of presentation".

[7] The paper can be accessed at http://dx.doi.org/10.1108/13673271311315196

Quality Attribute for Content	Min Value	Max Value	Average Value	Standard Deviation	Responses
Originality	4.00	91.00	57.59	23.83	27
Relationship to Literature review	10.00	90.00	63.44	21.11	27
Methodology	6.00	90.00	60.74	21.92	27
Results	4.00	84.00	60.52	22.18	27
Implications for research, practice and/or society	2.00	95.00	61.00	24.65	27
Quality of Communication & Language	0.00	100.00	74.74	22.68	27
Popularity of the subject	4.00	93.00	68.48	22.06	27
Appropriateness for the journal	2.00	94.00	70.67	24.10	27
Adequacy of literature review	8.00	92.00	64.67	23.57	27
Quality of research design	5.00	90.00	55.67	23.93	27
Adequacy of data analysis	10.00	86.00	51.67	24.10	27
Contributions to the literature	16.00	90.00	59.00	21.28	27
Legitimacy of conclusions	8.00	90.00	58.70	25.93	27
Practical/managerial significance	7.00	97.00	60.85	24.90	27
Clarity of presentation	0.00	100.00	73.26	25.57	27

Table 4-15 Statistics for Content Quality Assessment Attributes

The second list of attributes shown in Table 4-14 is associated with the format (schema) rating of the selected paper by the reviewers. Table 4-16 shows the statistics of such schema grading by the reviewers in this study.

Quality attributes for Schema	Min Value	Max Value	Average Value	Standard Deviation	Responses
Abstract	2.00	91.00	62.96	22.29	27
Introduction	6.00	90.00	62.00	21.97	27
Methodology	1.00	90.00	55.52	24.47	27
Results	3.00	91.00	55.63	24.87	27
Conclusion and Discussion	0.00	91.00	57.85	24.18	27
Adequate References	1.00	92.00	64.85	25.60	27
Accurate References	4.00	91.00	64.44	22.60	27
Up-to-date References	1.00	100.00	63.74	26.49	27

Table 4-16 Statistics for Schema Quality Assessment Attributes

Similar to the statistics from the content quality assessment, the opinions collected for the assessment of format (schema) from the 27 participants also ranges widely between 0 to 100, with relatively similar standard deviations, the lowest for the introduction and the highest for up-to date references. Finally, the overall grading and the summary of the final decisions for the selected manuscript are shown in Table 4-17 and Table 4-18 respectively.

	Min Value	Max Value	Average Value	Standard Deviation	Responses
Overall Grade	2.00	85.00	60.26	21.43	27

Table 4-17 Overall Quality Grading of Participants for Selected Manuscript

Final Decision	Response	%
Accept	2	7.5%
Minor Revision	16	59%
Major Revision	7	26%
Reject	2	7.5%
Total	**27**	**100%**

Table 4-18 Summary of Final Decision of Participants for Selected Manuscript

From Table 4-17 and Table 4-18, it can be concluded that in general, the manuscript used in this survey has been graded slightly above average with an overall grade of almost 60 out 100. However, the majority of the participants have declared their final decision as involving revision rather than direct acceptance or rejection.

Up to this section, the descriptive data that was collected during the survey has been summarised and discussed. This descriptive information provides a better picture for the decision analysis section on the background and preferences of the study group. In the next section, this study will focus on how decision patterns can be derived from such multidimensional observations to knowledge quality assessment, both in content and schema (knowledge resources).

4-4 Descriptive Analysis of Knowledge Quality Rating in Peer-Review

In the previous sections, the surveyed population preferences, decisions and quality ratings were discussed. From this section, based on the datasets collected from the survey, this study explores the analysis of these data and the decision patterns that can be extracted from both content and schema quality attributes that were discussed so far.

4-4-1 Correlation Comparison of Content Quality Rating Adopted for the Selected Publisher

In this part, based on the content quality assessment criteria that are adopted from the two publishers (IGI-Global and Emerald) and shown in Table 4-12 and 4-13, we test the correlation level between the content quality attributes from each publisher. Table 4-19 shows the correlation between the quality attributes from these two publishers using Pearson's method in SPSS. The correlation significance is at the 0.01 level for all values in Table 4-19 which shows a very good confidence level in the responses. The high correlation level between these variables shows the reviewer's generic approach toward assessing the selected manuscript. In other words, it means, that if a reviewer sees a manuscript as scientifically weak or strong, such observation will affect his or her quality assessment relative to the majority of assessment elements for the manuscript. This may suggest, in assessing the knowledge quality of the content of a manuscript, that the general understanding of a reviewer of a manuscript influences the whole quality assessment grading of the content.

IGI \ Emerald	Originality	Relationship to Literature	Methodology	Result	Implications for research, practice and/or society	Quality of Communication & Language
Popularity of the Subject	0.780	0.672	0.821	0.802	0.748	0.810
Appropriateness for the journal	0.737	0.590	0.702	0.808	0.767	0.843
Adequacy of literature review	0.717	0.908	0.769	0.828	0.865	0.776
Quality of Research design	0.829	0.648	0.796	0.883	0.787	0.803
Adequacy of data analysis	0.743	0.660	0.773	0.869	0.792	0.705
Contribution to the literature	0.799	0.659	0.755	0.850	0.819	0.733
Legitimacy of the conclusions	0.802	0.723	0.746	0.896	0.840	0.828
Practical /managerial significance	0.785	0.753	0.764	0.890	0.902	0.822
Clarity of presentation	0.787	0.745	0.733	0.858	0.823	0.963

Table 4-19 Correlation Table for Content Quality Attributes of Emerald and IGI

In other words, the reviewer's assessment grading is more subjective to his or her understanding of a manuscript rather than clearly differentiating the elements of the assessment; however, reviewers showed proper understanding in differentiating some of these elements. As the red circles show in Table 4-19, the highest values are associated with relatively similar quality attributes but with different wording from the two publishers. This can suggest, that the reviewers react similarly to different assessment forms that are relatively similar in assessment criteria but vary in terminologies.

High internal consistency (reliability) is also present in the quality attributes of each publisher (Cronbach's alpha of 0.966 for Emerald and 0.974 for IGI-global). This can be interpreted that a high level of correlation exists between the different quality attributes within the same form assessment by each publisher, meaning the quality attributes in each of the assessment sheets form a cohesive grading measurement for knowledge quality assessment.

4.4.2 Impact of Schema Quality Attributes in Overall Quality Grade

As mentioned earlier, while the participants were asked to evaluate the quality attributes for both content and schema separately, they were also requested to give an overall quality grade. In this part, in order to better visualize the impact of the schema (format) quality grading on overall quality grading, we have divided the overall quality into 5 categories: from 0 to 20 as "Very Poor", 20 to 40 as "Poor", 40 to 60 as "Fair", 60 to 80 as "Good and finally 80 to 100 as "Excellent". Figure 4-2 to Figure 4-9 shows the distribution of the overall grade corresponding to the schema (format) grade. As shown in bar legend colours, the blue, green, khaki, purple and yellow bars show the counts for Very Poor, Poor, Fair, Good and Excellent respectively.

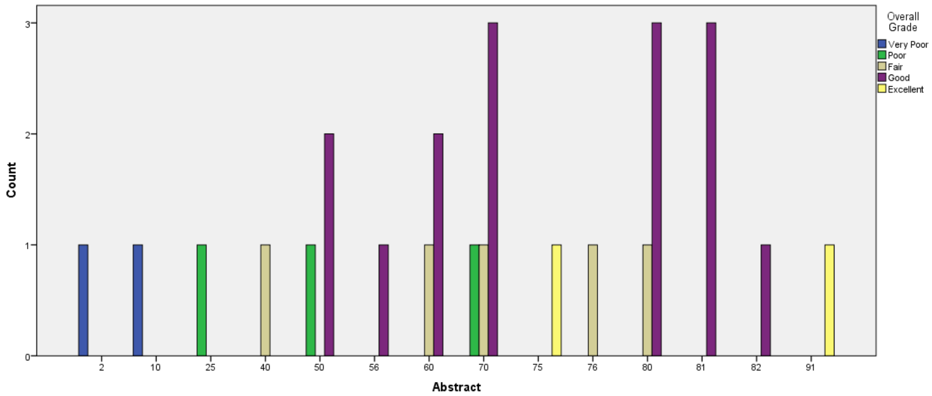


Figure 4-2 Abstract-Overall Grade Distribution

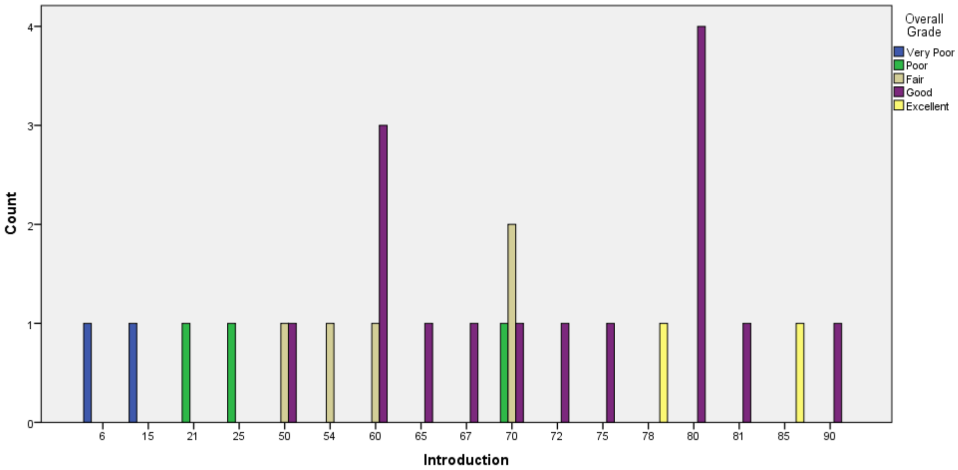


Figure 4-3 Introduction-Overall Grade Distribution

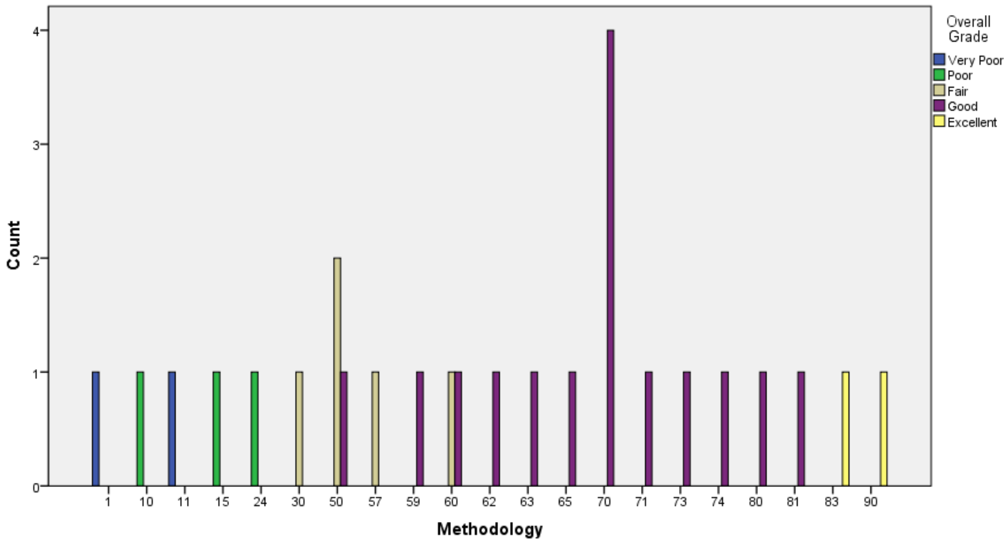


Figure 4-4 Methodology-Overall Grade Distribution

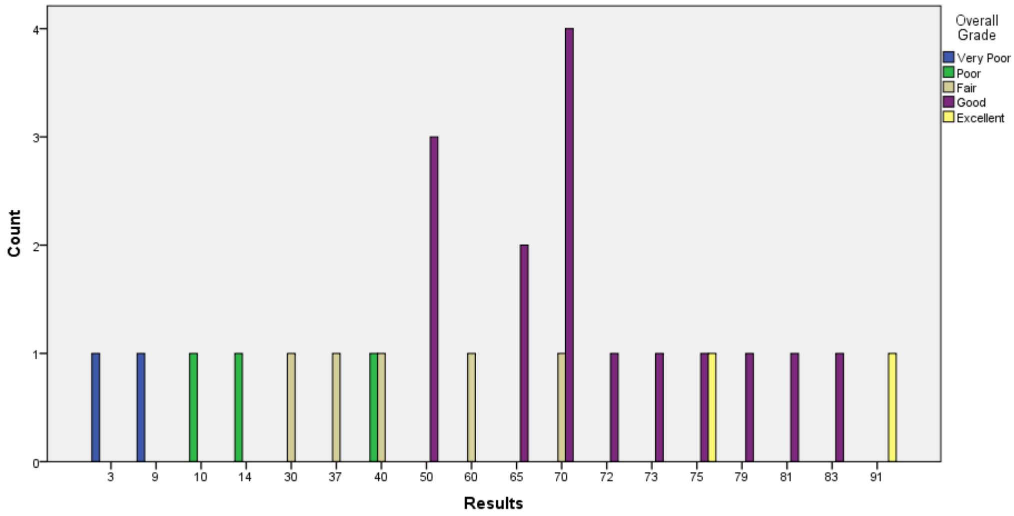


Figure 4-5 Result-Overall Grade Distribution

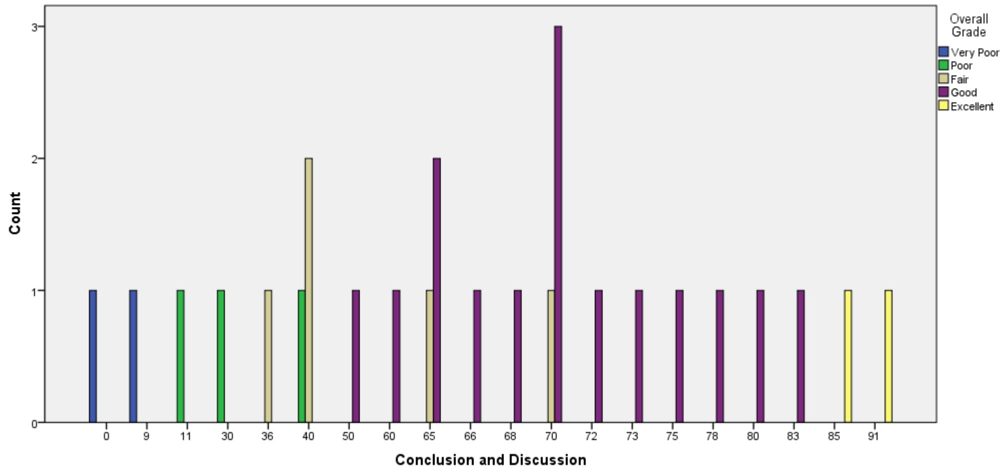


Figure 4-6 Conclusion & Discussion-Overall Grade Distribution

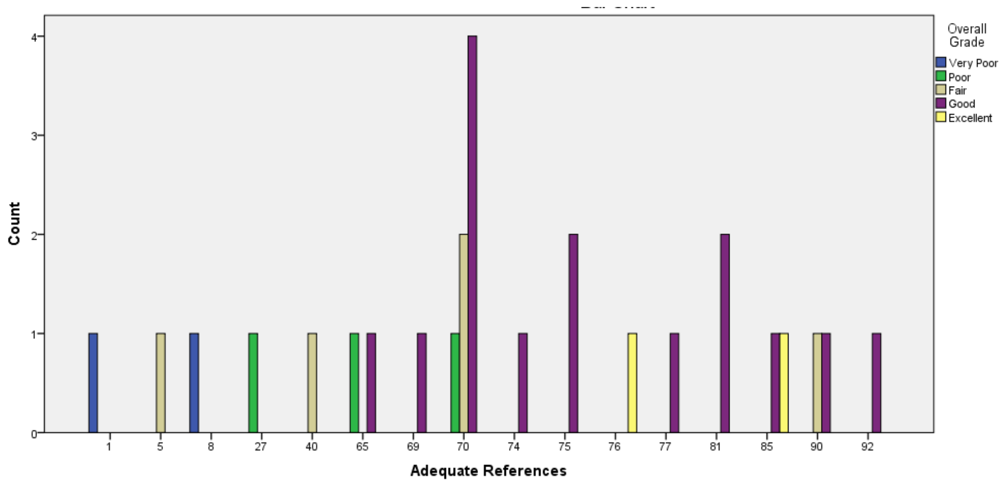


Figure 4-7 Adequate References-Overall Grade Distribution

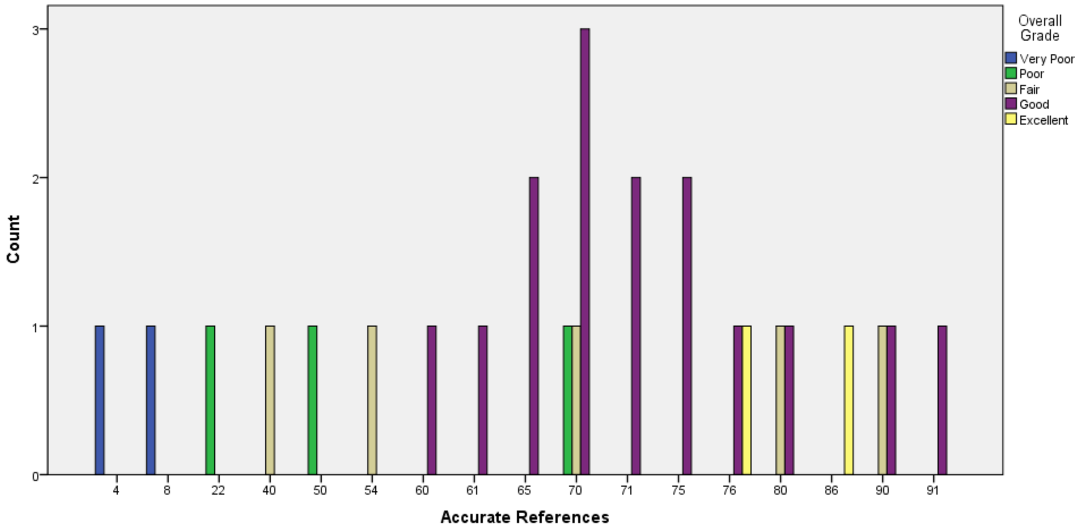


Figure 4-8 Accurate References-Overall Grade Distribution

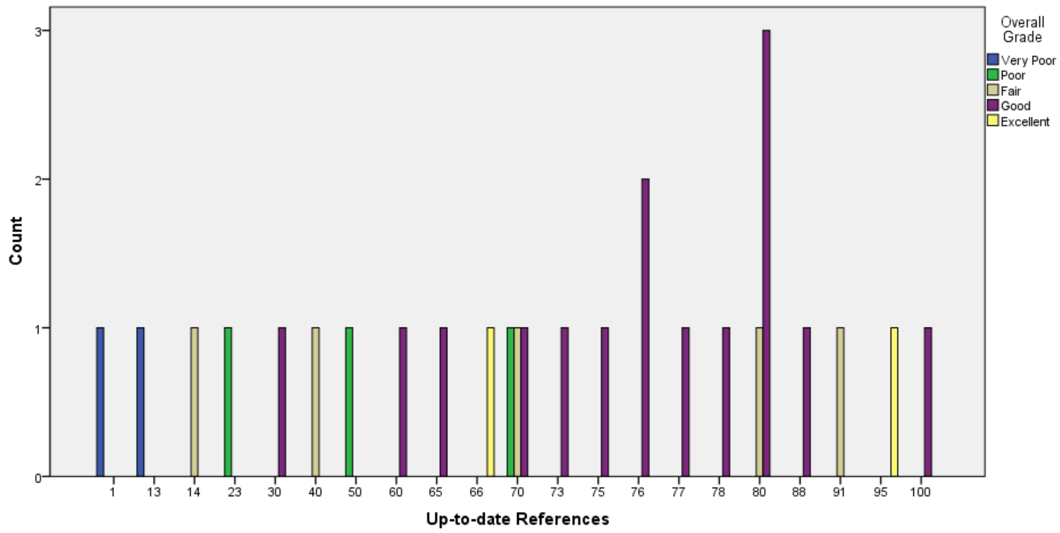


Figure 4-9 Up-to-date References-Overall Grade Distribution

From Figure 4-2 to Figure 4-9, the visualization depicted in the bar charts helps to better explore the role of each part of the manuscript format in a better way in relation to the overall grade given by the participants. Based on the colour and number of counts over the grading spectrum, the following patters can be observed:

1) *Abstract-Overall Grade Pattern*: The abstract format, when it is very poor, slides the overall grading to very poor for the manuscript. Only 1 count of poor abstract is associated with a good overall grade. In general, an overall grade of more than 50 (out of 100) has been achieved when the abstract has met at least a similar quality level in the general quality assessment of the manuscript.
2) *Introduction-Overall Grade Pattern:* The introduction grade is almost evenly distributed throughout the grading spectrum. The majority of all cases with an overall grade of more than 50 had an introduction of at least good. However, even poor and fair introductions were able to enjoy overall good grades (around 70) in 3 cases out of 27.
3) *Methodology-Overall Grade Pattern:* The methodology grade is relatively well correlated with the overall grading, being evenly distributed throughout the spectrum. A lower grade for methodology is associated with a lower overall grade, and only good to excellent methodology grades could take the overall grade into the good or excellent zone. This pattern shows the significance of the role that the structure of the methodology plays in assessment of the overall quality of the manuscript.
4) *Results-Overall Grade Patterns*: Similar to methodology, the results follow the methodology pattern with the exception that, in general, better results are expected for lower overall grades. In other words, good to excellent results are expected to take the overall grade to the fair to good level. It can be interpreted, in some cases, that the expected result could not achieve the same level of overall grade.

5) *Conclusions & Discussion-Overall Grade Pattern*: Conclusions and Discussion patterns are very similar to Results pattern in relation to the overall grade. A good result is expected in order to have at least a fair overall grade for the manuscript.

6) *Adequate references-Overall grade pattern:* This can be associated with the structure of the literature review in a manuscript. The diversity of opinion is relatively great for this parameter, as a poor or fair level of adequacy in references and literature review may result in good to excellent overall grades. This may suggest that, although many good to excellent papers are expected to have good adequacy of the references and literature review, it may not be a must in the reviewer's eyes in order to give the corresponding overall grade to the manuscript. The illustration in Figure 4-7 shows some of these exceptions.

7) *Accurate References-Overall grade Pattern*: From the illustration in Figure 4-8, accurate references are expected in majority of the cases if the paper is to have an overall grade of good to excellent. This shows the reviewer's expectation is to see the accuracy of the citation used in a manuscript for a relatively good overall grade.

8) *Up-to-date References-Overall Grade Pattern:* There is not a proper distribution pattern, as good or excellent up-to-date reference grades may have poor to fair overall grade and vice versa. While the majority of the cases with a good up-to-date references grade may still show a good overall grade, but no obvious pattern is evident from the illustration in Figure 4-9. Only very poor up-to-date references (in this case outdated references) in the view of the reviewer may be associated with poor overall grades.

4.4.3 Decision-Overall Grade Cross-tabulation

Here, we illustrate the relationship of the decision outcome with the overall grade that was obtained from the survey. Figure 4-10 shows the decision-overall grade distribution for the study survey using purple as reject, green as minor revision, khaki as major revision and blue as accept. As it can be seen, the decisions obtained from this study follow a natural pattern expected from a peer-review assessment. Nevertheless, the "Accept" decisions are associated with the overall grades associated with "Good", while "Excellent" overall grades (total count of 4) suggested minor revisions.

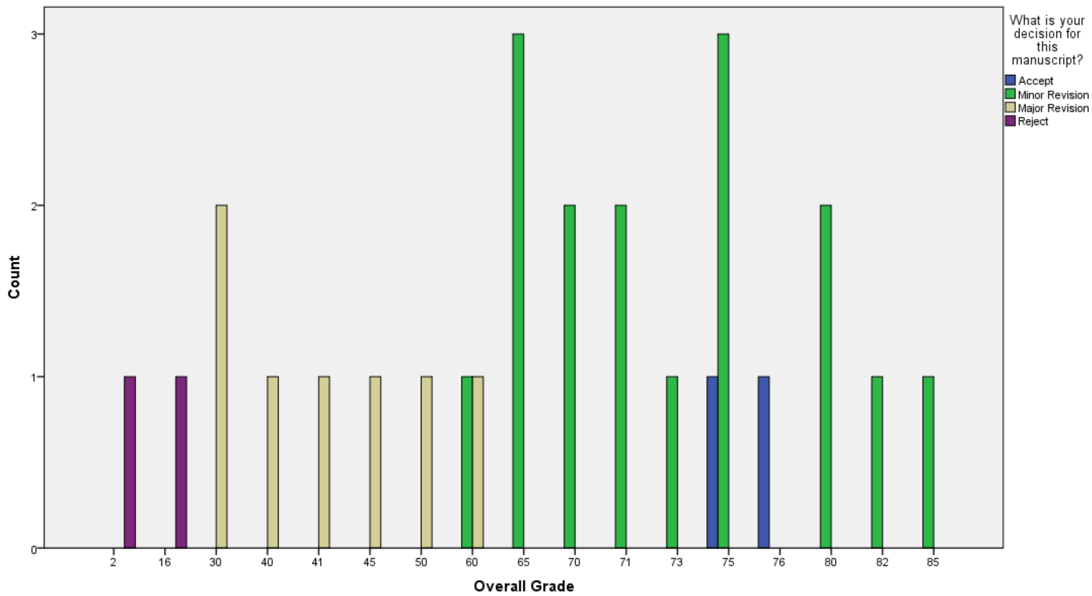


Figure 4-10 Decision-Overall Grade Pattern

We may now look to see where these decisions originate from, based on the background information that we have collected during the survey. Figure 4-11 shows the decision outcome and the academic position of the participants. As illustrated, the most diversity of decisions comes from participants with academic positions as professors. By taking into consideration Figure 4-12, it can be seen that the diversity in decision making comes from those who have either been in the peer-review process for many years or those who are quite new to this procedure.

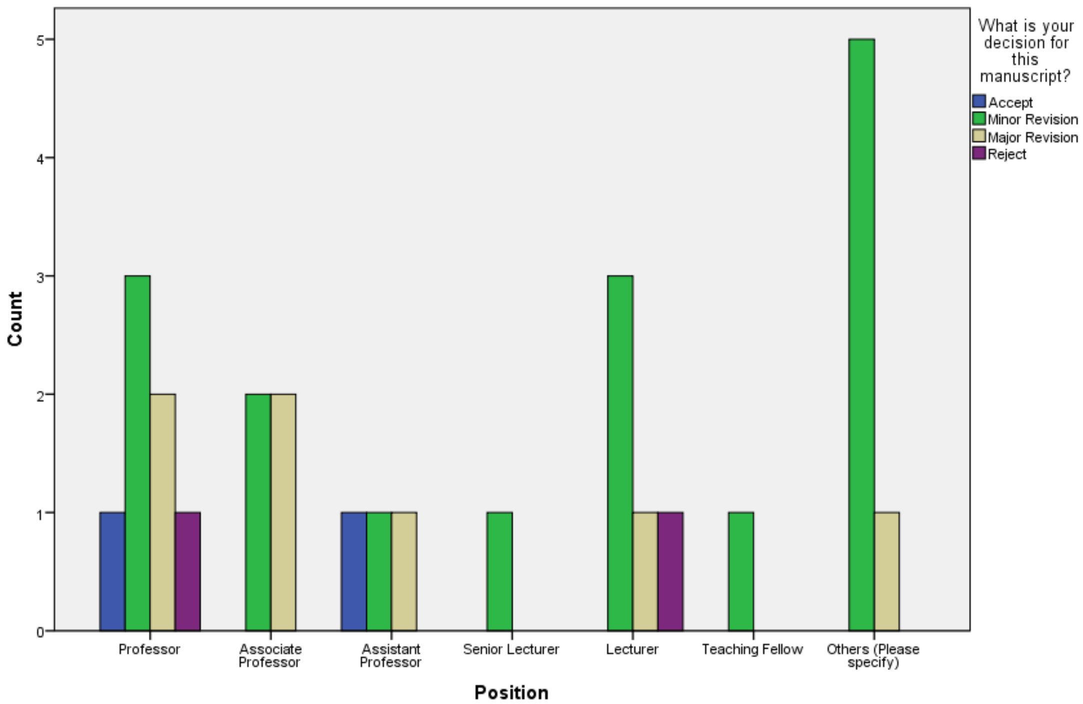


Figure 4-11 Academic Position-Decision Pattern

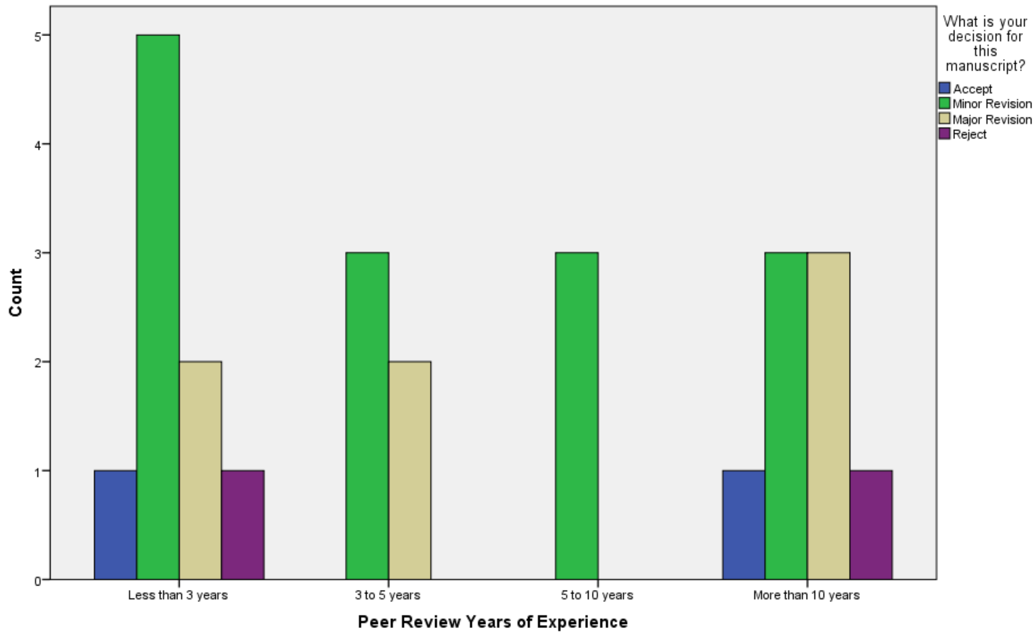


Figure 4-12 Peer Review Experience-Decision Patterns

This can be mapped in a way that the reliability of agreement in a decision is lower for senior and junior academic people in this study, while those with peer-review experience of 3 to 10 years have a more similar decision outputs.

Earlier in this chapter, the adopted journal list, with ranking in the KM discipline, were also shown (Table 4-5). In the survey, each participant was asked about their association as a reviewer or editorial member for the listed journals (both ranked and non-ranked). Table 4-20 is the summary of the reviewer's decisions associated with the KM journals ranking given in Table 4-5, adopted from Serenko & Bontis (2013).

Journal ranking	Counts for Decision-Journal Reviewer Association				
	Reject	Accept	Minor Revision	Major Revision	Total Counts
A+			2	2	4
A			4	3	7
B		1	7	8	16
C	1	1	2		4
Not Ranked	5			1	6

Table 4-20 Reviewer's Decision Associated with the KM domain Journals

In Table 4-20, the total count is based on the number of counts from journals of the same rank and not on individual reviewers. Apparently, one reviewer can be associated with several journals at the same time and in this case, the total number of count of 37 exceeds the total decisions of the 27 participants due to such multiple associations of reviewers to several journals. What can be concluded from Table 4-20 is, that reviewers who are not directly associated or familiar with the core publications of a discipline tend to reject more than others. On the other hand, those high ranking journals may tend to look for more perfection through multiple reviews for submitted manuscripts. There is one very important point to be noted here and that is that the open review process selected in this study reveals that work in the A+ journal of the KM discipline implies a level of trust in the content by those core contributors of a discipline. This effect can probably be investigated in a different study that evaluates the bias that originates from a reviewer's characteristics outside the peer-review process.

One of the other parameters that can be looked into is the core expertise of the reviewers in this study. Similar to the association with a journal, the reviewers can have expertise in different areas, thus the number of counts exceeds the number of individual decisions. As Table 4-20 shows, the diversity of the decisions varies with different backgrounds. Depending on the context and the research work being reviewed, the reviewer's expertise may influence the decision. A reviewer with some expertise may favour a content while another reviewer may dislike it. As Table 4-20 shows, in this study survey, participants with specific expertise assessed the manuscript differently. As the selected review paper was chosen from the KM domain and published in a KM journal, the percentage of decisions varies between different expertise backgrounds. For the participants of this study, the accumulated number of decisions based on expertise is listed in Table 4-20. While the true meaning of this table relies on a large samples size, however, even in the small sample in this survey, the variations can be seen as missing decisions in some domains of expertise. The decision diversity for the participants of this study shows that the decision diversity varies with the expertise backgrounds that were reported by the reviewers. Figure 14-3 shows this decision-expertise ratio for the expertise listed in Table 4-20. This may show different tendencies for decision making, especially when the expertise distances itself from the core knowledge domain (KM domain in this study)

Expertise	Counts for Decision-Expertise Association				
	Accept	Minor Revision	Major Revision	Reject	Total Counts
Knowledge Management	2	10	4	2	18
Intellectual Capital		2	4	1	7
Artificial Intelligence	1	3		1	5
Information Technology	1	1	1	1	4
Data Mining and Predictive Analysis	1	3		1	5
Semantic Web and Technologies		2		2	4
Computer Science and Software Engineering	1		1	2	4
Business and Management	1	6	2	2	11
Accounting and Finance			1	1	2
Economics				2	2
Intellectual Property				1	1
System Thinking and Modelling		3		2	5

Table 4-21 Reviewer’s Decision-Expertise Association

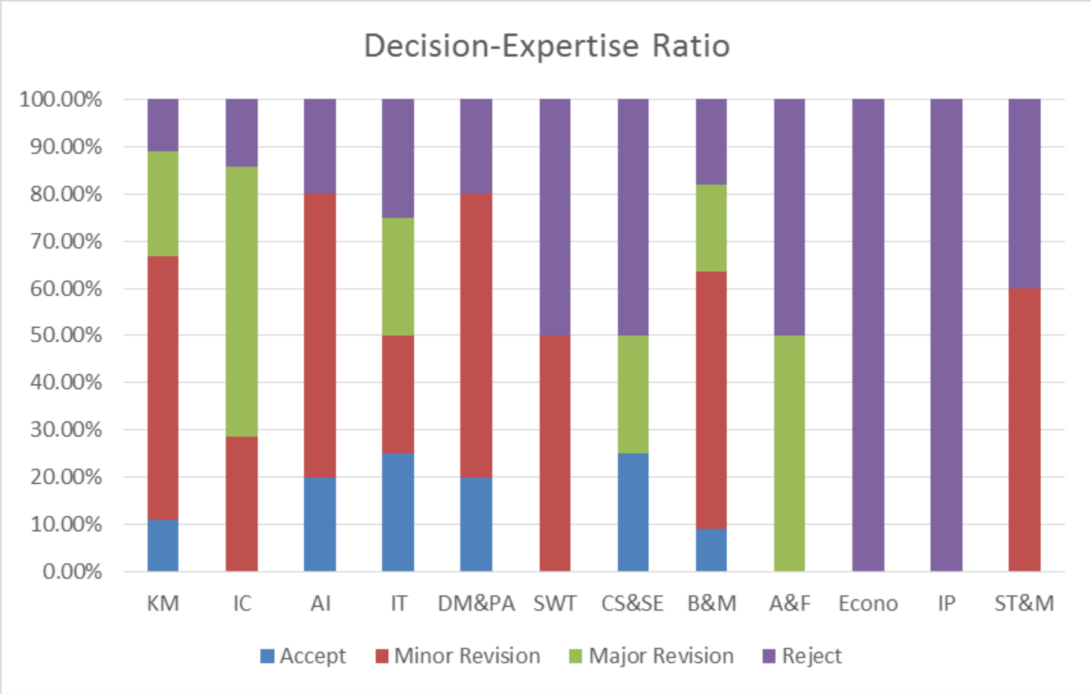


Figure 4-13 Decision-Expertise Domain Ratio of Participants

4-5 Peer Review Decision Pattern Analysis Using Decision Trees

In the previous section, we discussed the descriptive analysis of the data collected from the survey. The descriptive data was mainly about the summary of the data directly collected throughout the survey, and an in-depth analysis of the correlation and cross tabulation of data in the latter part of the previous section. In this section, we explore the decision patterns that originate from the reviewer preferences and quality attributes both for the content and schema. The decision trees provide an insight into the priorities each quality attribute may have in making a decision for a manuscript, and how these knowledge quality attributes are branched to create the decision path for a manuscript.

The decision tree is defined as a classification method that "Partitions the data into smaller subsets where each subset contains (mostly) responses of one class". (Kotu &

Deshpande, 2015). There is no restriction on the variable types as input for decision trees. While decision trees, in some cases (e.g. small datasets), tend to over-fit the data so that small changes in input may have a great impact on the result, nevertheless they are very suitable for clustering and predictions without the need for any further normalization of the data and allow very easy interpretation. In this study, we use the SPSS decision tree classifier to analyze the decision patterns of derived from the reviewer's preferences and quality attributes. The objective is the segmentation based on the decisions. In this study, the target variable is the decision (accept, reject, minor revision or major revision) and the variable can be either a reviewer's preference rankings or a reviewer's knowledge quality content and schema grades. There are basically three types of data in SPSS 1- *Nominal*: when a variable category represent a category, 2- *Ordinal*: when the value of a variable has some kind of intrinsic ranking (levels) and finally, 3- *Scale*: in which the value of the variable represents order categories on a continuous scale with some meaningful metrics (age, distance etc.). In this study, we have all the three types for analysis. The decision variable is the nominal value, as we have 4 decisions of accept, reject, minor revision and major revision in which each can be considered as a category of decision. We also have ranks for preferences as ordinal, where the lowest number (1) shows the highest (first) priority. We also have quality grading for content and schema which is a scale variable ranging between 0 to 100.

Decision trees also have different growth methods. Among many different growth methods, there are three popular methods that can be used in SPSS. 1-*CHAID (Chi-square Automatic Interaction Detection:* which at every step, it chooses the predictor (independent) variable with the strongest interaction with the dependent variable. There is also *Exhaustive CHAID* which tests all the possible splits for each independent variable 2- *CRT or Classification and Regression Trees:* that splits data into homogenous segments with respect to the dependent variable and finally *3- QUEST or Quick, Unbiased Efficient Statistical Tree*: that is a more recent approach to decision trees and has less bias in comparison with the other methods when there are many predictors with many categories. The QUEST method can be used when the target (dependent variable) is only nominal (similar to decisions in this study).

Figure 4-14 shows the setup box for decision variables in decision trees. While the dependent variable (Reviewer's Decision) stays the same throughout the test, the independent variable changes based on the types of quality attributes and reviewer preferences.

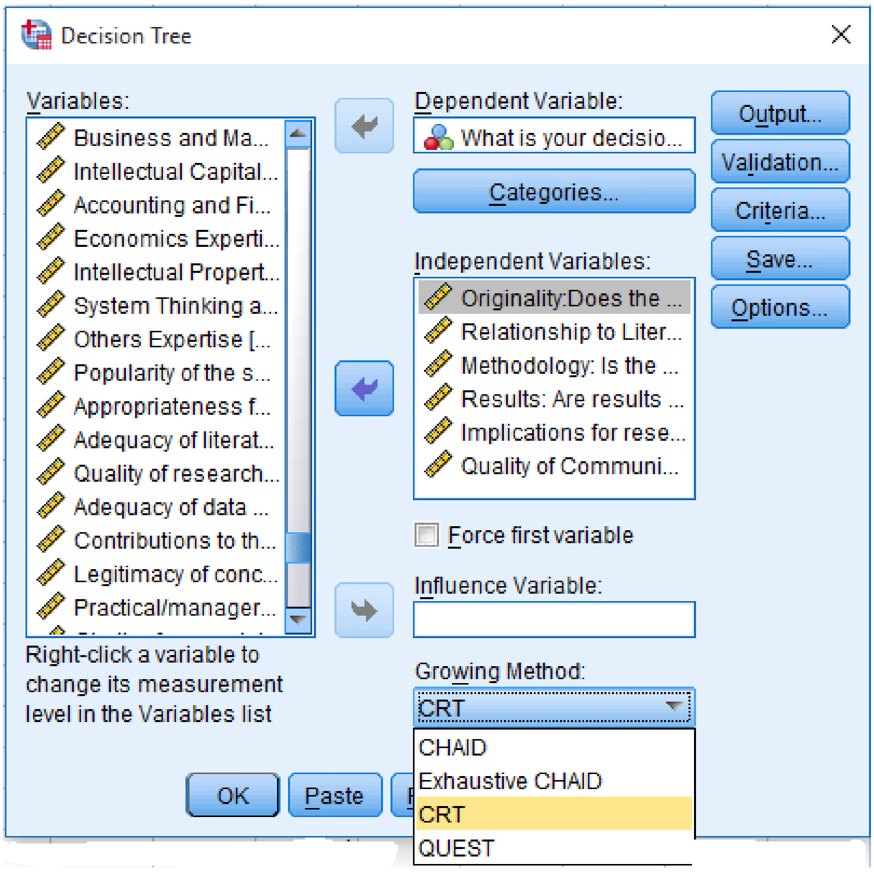


Figure 4-14 Decision Tree Setup Box

With this introduction to decision trees, we explore the decision patterns based on the preference and quality inputs from the survey. We use the three growth methods of Exhaustive CHAID, CRT and QUEST for each scenarios and will proceed with the details of the method that has more reliability (less risk). Moreover, as the sample size is relatively small (27 participants), we keep our minimum number of cases (which is 1) for both parent node and child node. This means, as long as at least 1 case that can be categorized within a node, the decision tree can continue branching in and out of that node.

This study also checks the performance of decision trees obtained from the dataset analysis based on the best growth method (the growth method with the lowest risk) achieved in each scenario. It should also be noted that, not all the performance and validation measures are available under different growth methods or different types of target values. Hence, the validation and performance assessment is based on the best performing growth method obtained from the decision tree structure.

With the target value of decision as the categorical dependent variable (nominal), there are some node performance characteristics that can be tested. One of the evaluation characteristics that will be tested in this study is the *Gains* for nodes. The node's gain percentage is the percentage of the total number of cases in the *Terminal Node*. The terminal node is the node at the end of the tree branch, with no child node. In other words, the gain percentage shows the performance of the decision tree in classification of the target value for a specific target, which is either accept, reject, major revision or minor revision in this study.

There are two major tests for performance and validation in SPSS. 1) *Split-Sample validation:* which creates a training sample and a test sample based on the percentage of the total sample size. For the performance and validation test in this study, 20%, 50% and 80% are chosen as the percentages of sample size training, which means, dividing the sample size into two subsets, one for training and one for testing. 2) *Crossvalidation*: which divides the sample size into subsets or into different folds. As both types of validation manipulate the size of the dataset, these methods are further explored as the underlying assessment factor for comparison of public open review and peer review in chapter 5.

4.5.1 Decision Trees Based on Content Quality Attributes Adopted from Emerald

The first decision tree is built on the six content quality attributes adopted from Emerald, and was used in the survey. These six factors are the ones shown in Table 4-12 1-Originality 2-Relationship to Literature 3-Methodology 4-Result 5-Implications for research, practice and/or society and 6-Quality of Communication & Language. As mentioned, we run three growth methods (exhaustive CHAID, CRT and QUEST) for each scenario in order to find the one which provides more coverage (less risk) for categorization of the independent variables. For each scenario, we show how the comparison is done and only proceed with the growth method with the lowest risk for decision tree structures.

For this first scenario, the six factors mentioned above are the independent variable and the decision variable is dependent in all case (Figure 4-14). After running the decision tree analysis, a summary of the analysis is shown in a tabular format. Table 4-22 to Table 4-24 show the differences in risk level using these three growth methods in this scenario.

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	2	0	0	0	100.0%
Minor Revision	1	15	0	0	93.8%
Major Revision	0	2	5	0	71.4%
Reject	0	0	0	2	100.0%
Overall Percentage	11.1%	63.0%	18.5%	7.4%	88.9%

Growing Method: **EXHAUSTIVE CHAID**

Risk

Estimate	Std. Error
.111	.060

Table 4-22 Risk Assessment Using exhaustive CHAID Growth Method for Emerald

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	2	0	0	0	100.0%
Minor Revision	0	16	0	0	100.0%
Major Revision	0	0	7	0	100.0%
Reject	0	0	0	2	100.0%
Overall Percentage	7.4%	59.3%	25.9%	7.4%	100.0%

Growing Method: **CRT**

Risk

Estimate	Std. Error
.000	.000

Table 4-23 Risk Assessment Using CRT Growth Method for Emerald

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	0	2	0	0	0.0%
Minor Revision	0	15	1	0	93.8%
Major Revision	0	1	6	0	85.7%
Reject	0	0	2	0	0.0%
Overall Percentage	0.0%	66.7%	33.3%	0.0%	77.8%

Growing Method: **QUEST**

Risk

Estimate	Std. Error
.222	.080

Table 4-24 Risk Assessment Using QUEST Growth Method for Emerald

As can be seen, based on the analysis using the above three methods, the lowest risk in this scenarios is for the CRT growth method. The exhaustive CHATS risk is at 0.111, meaning 1-0.111=0.889*100 =88.9%, which is similar to the overall percentage in the classification table. Similarly, the QUEST method risk level stands at 0.222 which only covers 77.8% of the overall cases. The CRT, with a risk level of 0 in this scenario covers 100% of the cases in the dataset. In other words, not only has CRT the lowest risk level in this scenario, but its risk level of 0 means that all the cases in the dataset are covered. Thus, based on this risk assessment, this scenario will proceed with the decision tree obtained with the CRT method.

Figure 4-15 shows the decision tree obtain from the CRT growth for the content quality attributes adopted from Emerald.

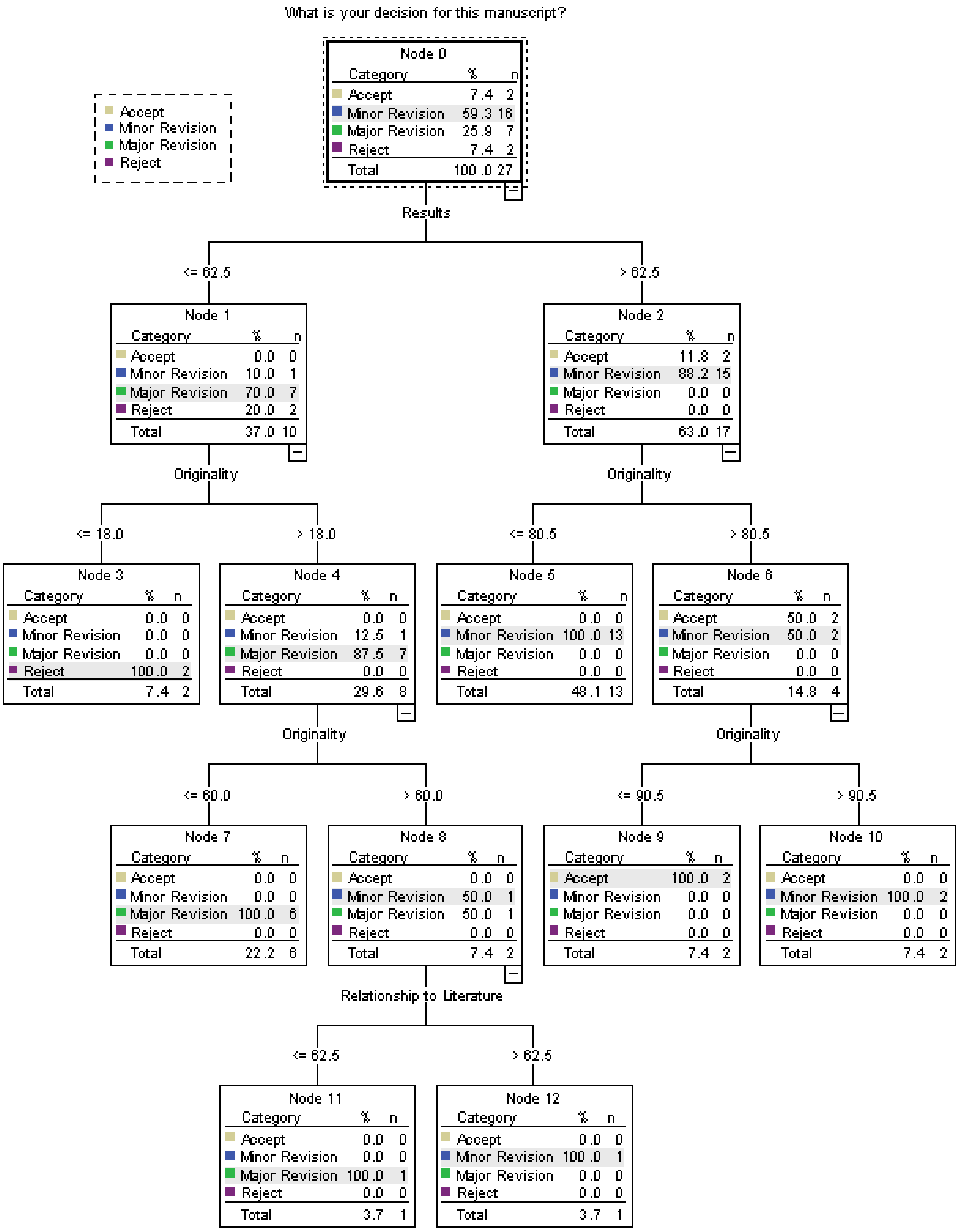


Figure 4-15 Decision Tree Using CRT for Content Quality Attributes Adopted from Emerald

From the decision tree shown in Figure 4-15, out of 6 quality content attributes, three (Results, Originality and Relationship to Literature Review) can predict 100% of the decisions made in our dataset from the survey. The highlighted decision in each node is the best predicted output of that node. In this scenario, as can be seen in Figure 4-15, the decision tree can be interpreted as follows:

1- The grade for "Results" is the best predictor for the final decision. If the grade for results of the selected manuscript is lower than 62.5, which is in the Good range according to the survey categorization (Refer to the legend in Figure 4-1), 70% of the reviewers are expected to require the manuscript to have major revision, 10% of the reviewers may request minor revision and 20% of the reviewers may reject the paper.
2- For grades of "Results" more than 62.5 (out of 100), it is expected that the manuscript won't be rejected or require minor revision. Nevertheless, 88.2% of the reviewers may request minor revision while 11.8% may directly accept the paper.
3- For those manuscripts with the condition of step 1, if the "Originality" grade is over 80.5 (the excellent category), then it is expected that 50% of the reviewers accept the manuscript while 50% may request minor revisions.
4- For those manuscripts with the condition of step 2 and with "Originality" grade below 18 (in the "Very Poor" category), it is expected that all will be rejected while for those over 18, it is expected that majority of reviewers, 87.5% will request major revisions while 12.5% of reviewers may ask for minor revisions.

The above interpretation continues similarly throughout the decision tree till the last node (the node at the bottom).

As we showed in the descriptive data earlier in this chapter, there is a medium to high correlation in all quality assessment variables between the two publishers. Here, based on the CRT method with the lowest risk method in this scenario, we can see, with almost 50% of the content quality attributes from the same publisher (three out six), the whole dataset can be mapped into the decision tree. This suggests that some of the quality attributes in the dataset can be predicted by another quality attribute with relatively similar importance. Figure 4-16 shows the normalized importance of these content quality attributes. As it can be seen, in the case of the content quality attributes used in Emerald's review forms, in general, the major differences can be categorized into two groups, meaning, the quality content of any group can predict the decision tree without the need of other content quality attributes.

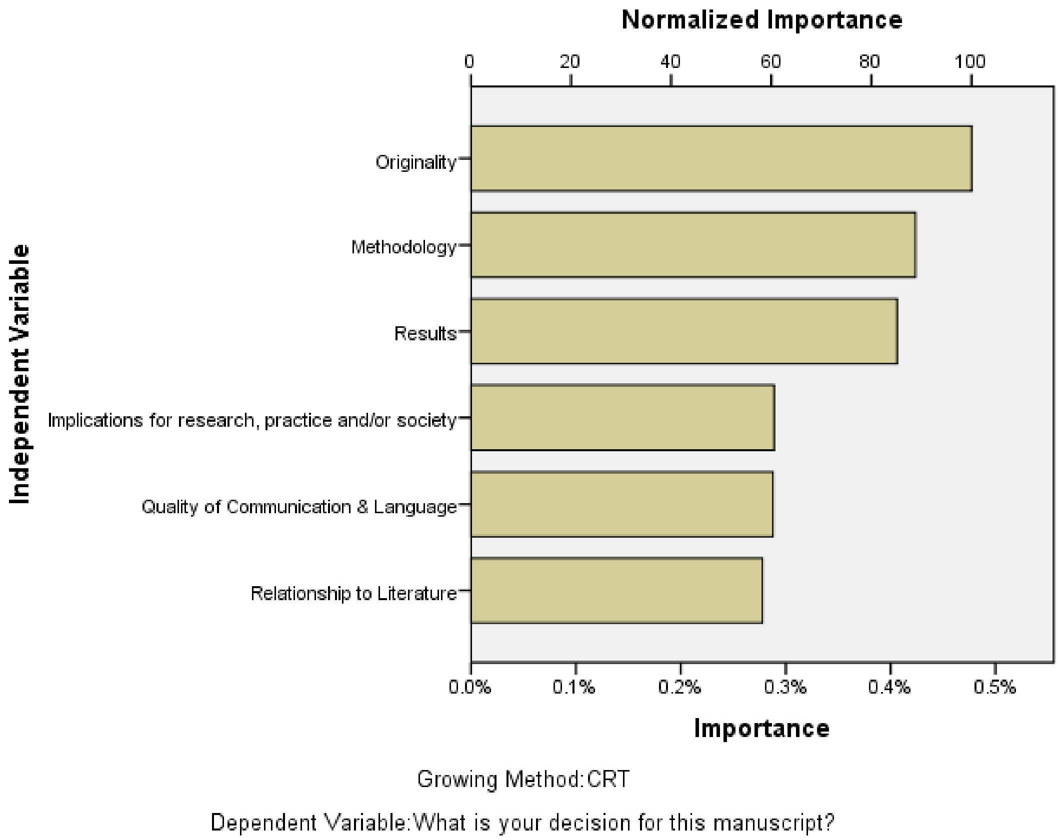


Figure 4-16 Normalized Importance of Quality Variables Adopted from Emerald

This, 50% or three attributes out of six, could enable construction the whole decision tree map.

In order to better understand the performance of the decision tree obtained from the content quality attributes adopted from Emerald, we examine the gain percentage for each of the target decisions. Figure 4-17 shows the node's gain performance for each of the target values obtained from the decision tree shown in Figure 4-15.

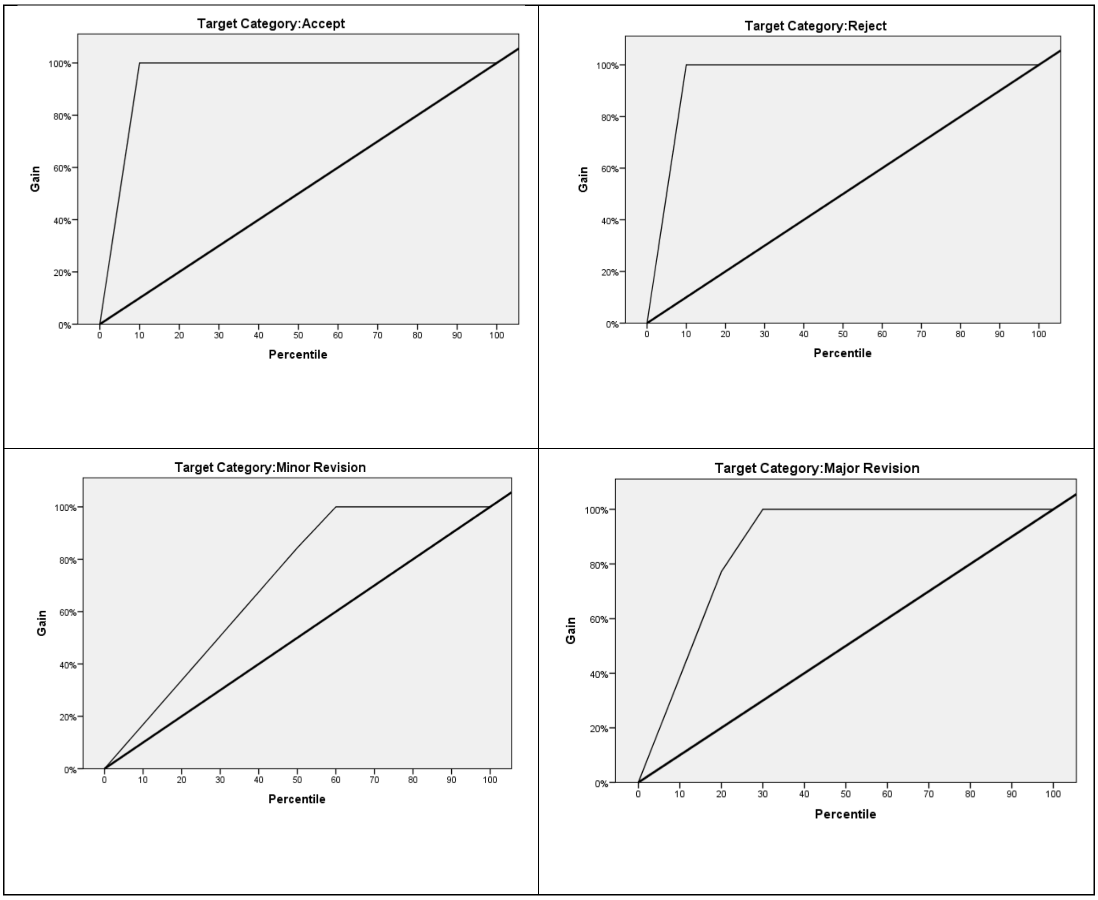


Figure 4-17 Gain Percentage for Decision Tree of Content Quality Attributes adopted form Emerald

As figure 4-17 shows, with the decision tree of the dataset in this scenario, we expect to capture both the accept or reject decisions from less than 10% of the sample size while, for major revisions, we may need to know around 25% of the sample size. To know about the decision on minor revision, we need to have around 55% of the sample size to be sure about the outcome. This suggests that arriving at accepting or rejecting a paper is usually more straightforward than a request for revision of the paper and the uncertainty about requests on minor revision or major revisions.

The summary of the decision tree for all types of decisions in this scenario is shown in Figure 4-18.

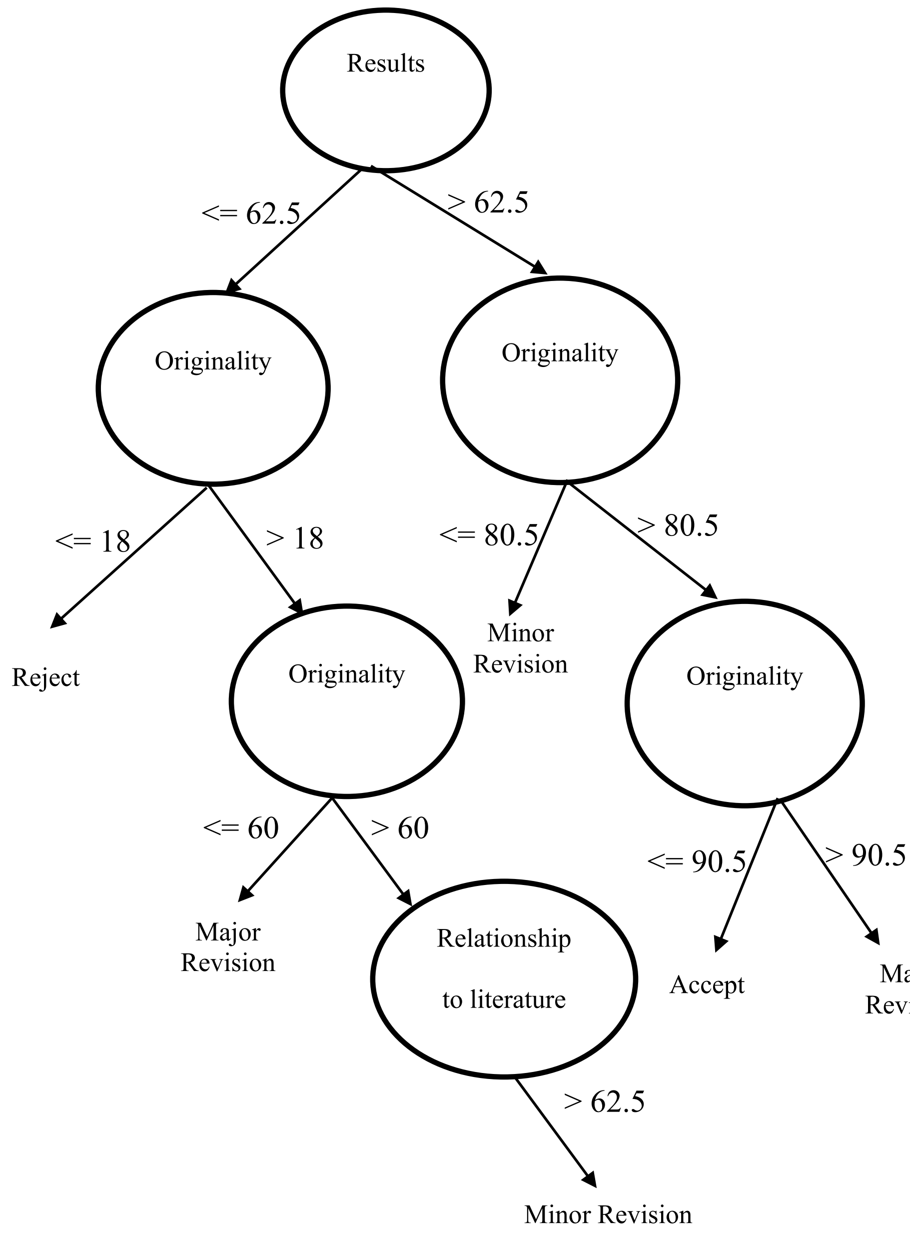


Figure 4-18 Decision Path for Content Quality Attributes Adopted from Emerald

4.5.2 Decision Trees based on Content Quality Attributes adopted from IGI-Global

The second scenario is the dataset acquired from the content quality attributes adopted from IGI-global as the other publisher. As shown in Table 4-14 earlier in this chapter, there are totally 9 content quality attributes for this scenario: 1- Popularity of the Subject 2-Appropriateness for the journal 3-Adequacy of literature review 4-Quality of research design 5-Adequacy of data analysis 6-Contribution to the literature 7-Legitimacy of the conclusions 8-Practical/managerial significance and 9-Clarity of presentation.

Similar to scenario one, we run the three growth methods of exhaustive CHAID, CRT and QUEST for the risk assessment level of each method in this scenario. Table 4-25 to Table 4-27 shows the risk level of each growth method for this scenario. As the risk assessment shows, similar to scenario one, the CRT risk level is 0, meaning it can cover 100% of the cases in the dataset.

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	0	2	0	0	0.0%
Minor Revision	0	15	1	0	93.8%
Major Revision	0	0	7	0	100.0%
Reject	0	0	0	2	100.0%
Overall Percentage	0.0%	63.0%	29.6%	7.4%	88.9%

Growing Method: **EXHAUSTIVE CHAID**

Risk

Estimate	Std. Error
.111	.060

Table 4-25 Risk Assessment Using exhaustive CHAID Growth Method for IGI

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	2	0	0	0	100.0%
Minor Revision	0	16	0	0	100.0%
Major Revision	0	0	7	0	100.0%
Reject	0	0	0	2	100.0%
Overall Percentage	7.4%	59.3%	25.9%	7.4%	100.0%

Growing Method: **CRT**

Risk

Estimate	Std. Error
.000	.000

Table 4-26 Risk Assessment Using CRT Growth Method for IGI

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	0	2	0	0	0.0%
Minor Revision	0	16	0	0	100.0%
Major Revision	0	0	7	0	100.0%
Reject	0	0	0	2	100.0%
Overall Percentage	0.0%	66.7%	25.9%	7.4%	92.6%

Growing Method: QUEST

Risk

Estimate	Std. Error
.074	.050

Table 4-27 Risk Assessment Using QUEST Growth Method for IGI

By having the CRT as the lowest risk growth method, the analysis for this scenario proceeds with the decision tree based on CRT. Figure 4-19 shows the decision tree obtained for this scenario, based on the CRT growth.

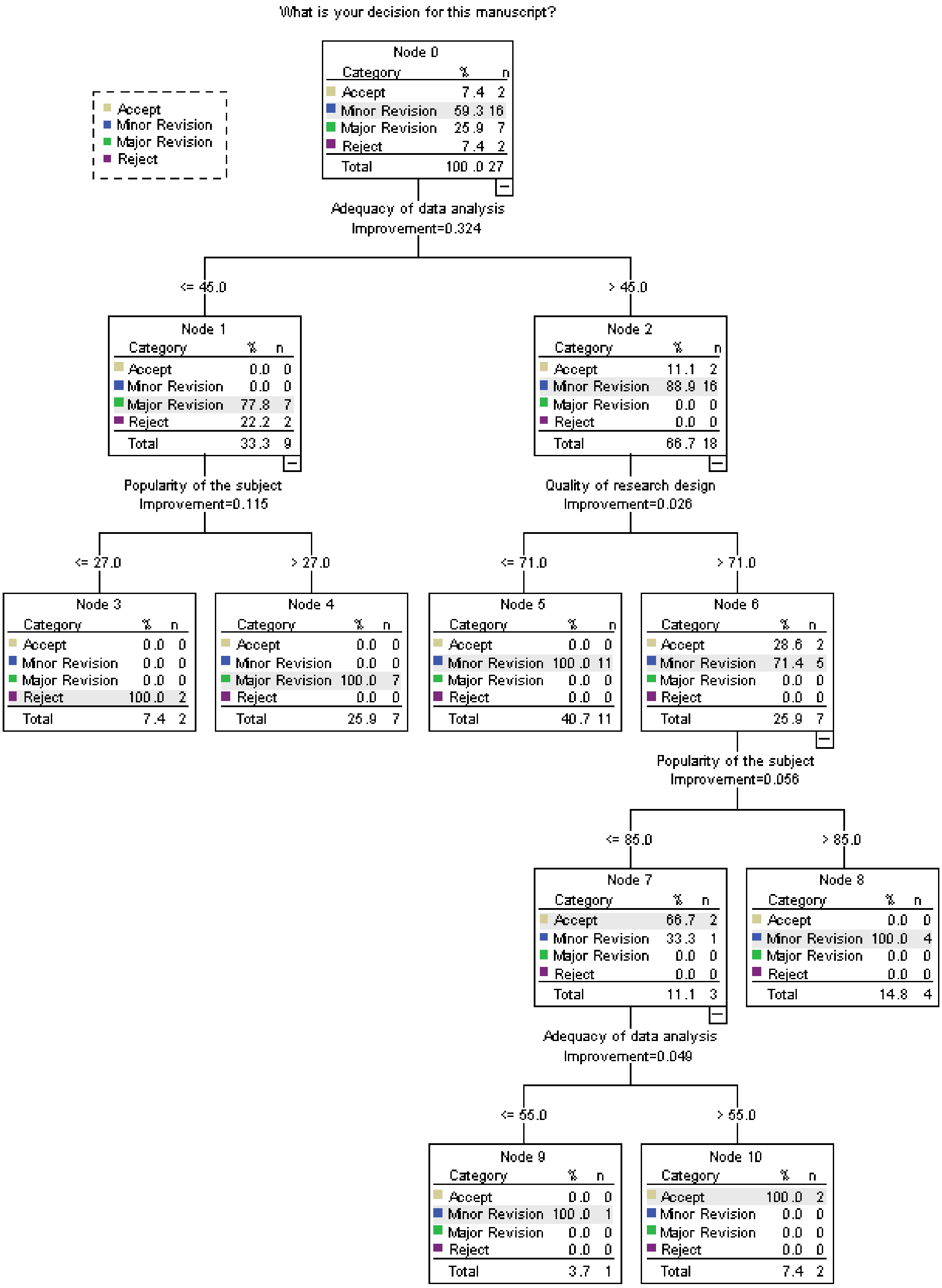


Figure 4-19 Decision Tree Using CRT for Content Quality Attributes Adopted from IGI

As it can be seen in Figure 4-19, out of nine content quality attributes adopted from IGI-Global, with three attributes of adequacy of data analysis, popularity of the subject and quality of research design, 100% of the cases can be classified for the dataset obtained from the survey. The interpretation for the decision tree in Figure 4-19 can be summarised as follow:

1- The grade for "Adequacy of data analysis" is the best predictor for the final decision. If the grade for the result of the selected manuscript is lower than 45, which is in the Fair range according to the survey categorization (refer to the legend in Figure 4-1), 77.8% of the reviewers are expected to request the manuscript have major revision and 22% of the reviewers may reject the paper.
2- For grades of "Adequacy of data analysis" more than 45, it is expected that the manuscript won't receive rejection but there is a very high chance of minor revision (88.9%). Nevertheless, 11.1% may accept the manuscript at this stage.
3- It is the grade of "Quality of research design" that can take the paper to the acceptance zone, or request for minor revisions (node 5 and node 6)
4- The grades for "Popularity of research design" and "adequacy of data analysis" are recursive variables that end the decision tree in regard to decisions on acceptance or minor revisions of the manuscript.

The above interpretation can be similarly expanded to all decision tree branches, based on the predicting variable for each node.

While in the previous scenarios, we had one recursive quality variable (originality) that was repealed most in the decision tree, in this scenario, we have two variables "popularity of the subject" and "adequacy of the data analysis". Again, this shows the high level of correlation among the content quality assessment variables where, having three variables can predict all the cases obtained from the adopted nine variables. Figure 4-20 shows the normalized importance of all these nine variables with respect to each other.

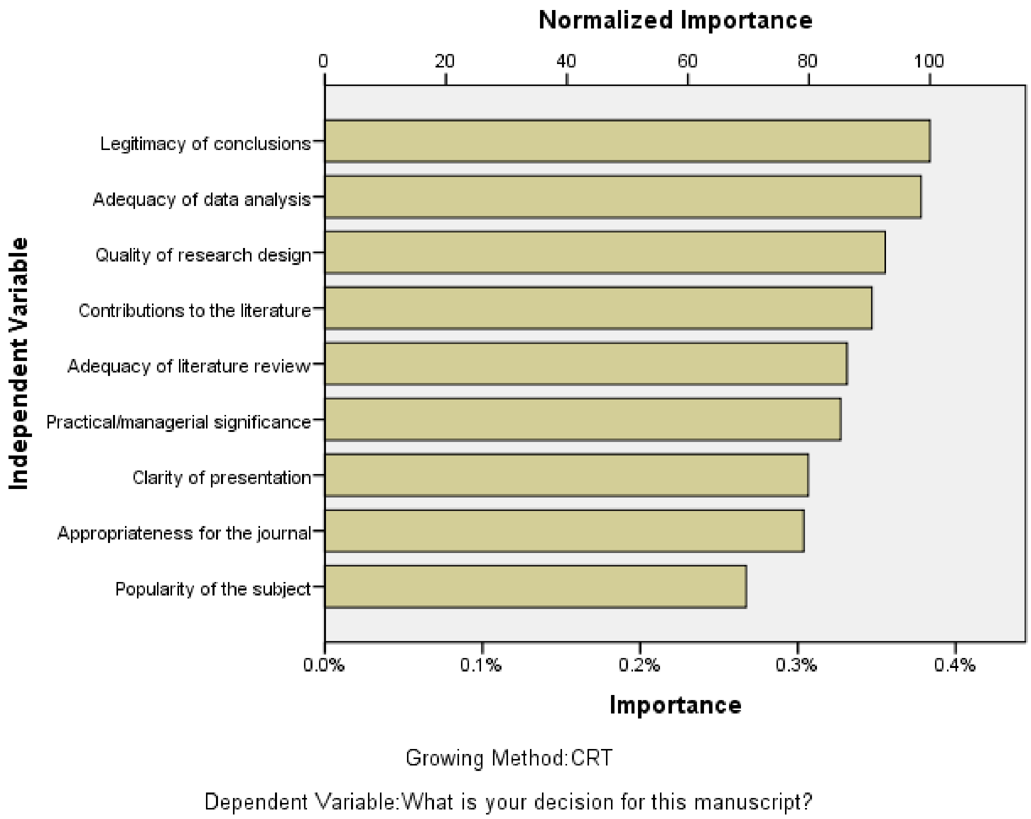


Figure 4-20 Normalized Importance of Content Quality Variables Adopted from IGI

In order to check the performance of these attributes on the target decision, we check the gain percentage of each decision as the result of the attributes appearing in the decision tree. Figure 4-21 shows the gain percentage in this scenario. As it can be seen, the gain patterns of scenario two (IGI) is relatively identical to scenario one (Emerald), with both acceptance and rejection reached with having less than 10% of the sample size, while major revision reaches 100% with around 25 of the samples and the slowest is the minor revision that reaches 100% at around 56%.

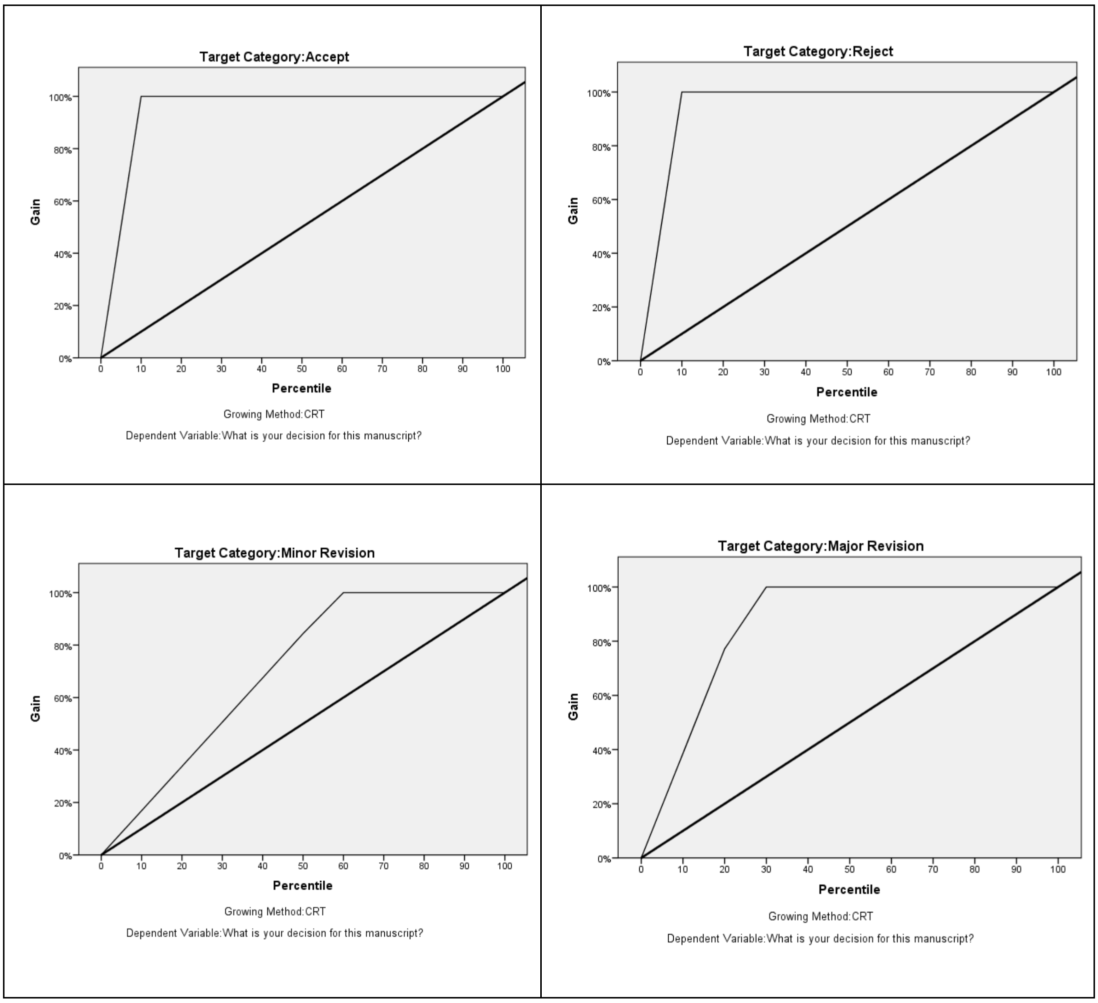


Figure 4-21 Gain Percentage for Decision Tree of Content Quality Attributes adopted form IGI

Comparing Figure 4-17 and Figure 4-19, it can be concluded that both series of attributes adopted from Emerald or IGI can predict with almost similar performance for the target decisions.

A summary of the decision tree for the terminal node, which is the decision path for this scenario, is shown in Figure 4-22.

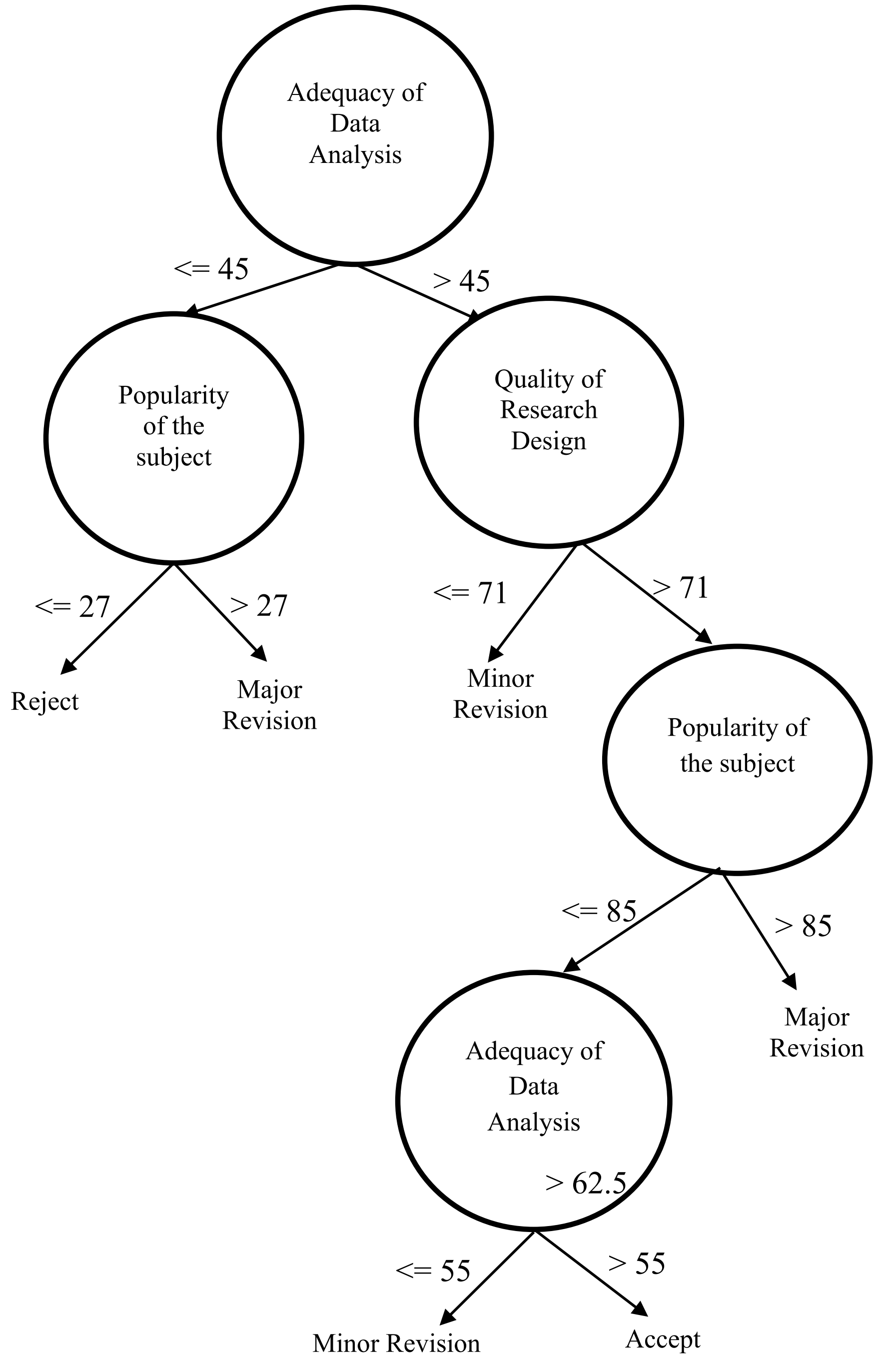


Figure 4-22 Decision Path for Content Quality Attributes Adopted from IGI

4.5.3 Decision Trees based on Schema Quality Attributes

In the previous parts, we discussed the content quality attributes that were adopted from two major publishers in the scientific domain of this study in knowledge management. In this part, the dataset collected from the survey on the schema grading of the selected manuscript is discussed. As mentioned earlier, this schema follows the most common formats for manuscripts, and are related more to the structure of the manuscript rather than the content. In other words, these are the factors that come into focus on the initial screening of the manuscript and skimming through the content. While the dependent variable stays the same as the decision on the manuscript, the independent variables in this scenario are the ones shown in table 4-14. These eight variables are 1- Abstract 2- Introduction 3- Methodology 4- Conclusion and Discussion 5- Results 6- Adequate References 7- Accurate References 8- Up-to-date References. In the first step, we try to assess the most suitable growth method for this scenario. Tables 4-28 to 4-30 show the risk level from each growth method for this scenario.

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	0	2	0	0	0.0%
Minor Revision	0	16	0	0	100.0%
Major Revision	0	1	6	0	85.7%
Reject	0	0	0	2	100.0%
Overall Percentage	0.0%	70.4%	22.2%	7.4%	88.9%

Growing Method: **EXHAUSTIVE CHAID**

Risk

Estimate	Std. Error
.111	.060

Table 4-28 Risk Assessment Using exhaustive CHAID Growth Method for Schema Quality Attributes

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	2	0	0	0	100.0%
Minor Revision	0	16	0	0	100.0%
Major Revision	0	0	7	0	100.0%
Reject	0	0	0	2	100.0%
Overall Percentage	7.4%	59.3%	25.9%	7.4%	100.0%

Growing Method: **CRT**

Risk

Estimate	Std. Error
.000	.000

Table 4-29 Risk Assessment Using CRT Growth Method for Schema Quality Attributes

Classification

Observed	Predicted				
	Accept	Minor Revision	Major Revision	Reject	Percent Correct
Accept	0	2	0	0	0.0%
Minor Revision	0	15	1	0	93.8%
Major Revision	0	1	6	0	85.7%
Reject	0	0	2	0	0.0%
Overall Percentage	0.0%	66.7%	33.3%	0.0%	77.8%

Growing Method: **QUEST**

Risk

Estimate	Std. Error
.222	.080

Table 4-30 Risk Assessment Using CRT Growth Method for Schema Quality Attributes

Similar to the two previous scenarios, the CRT growth method has the lowest risk and covers 100% of the case in this scenario. Figure 4-23 shows the decision tree based on the eight independent variables of the schema assessment.

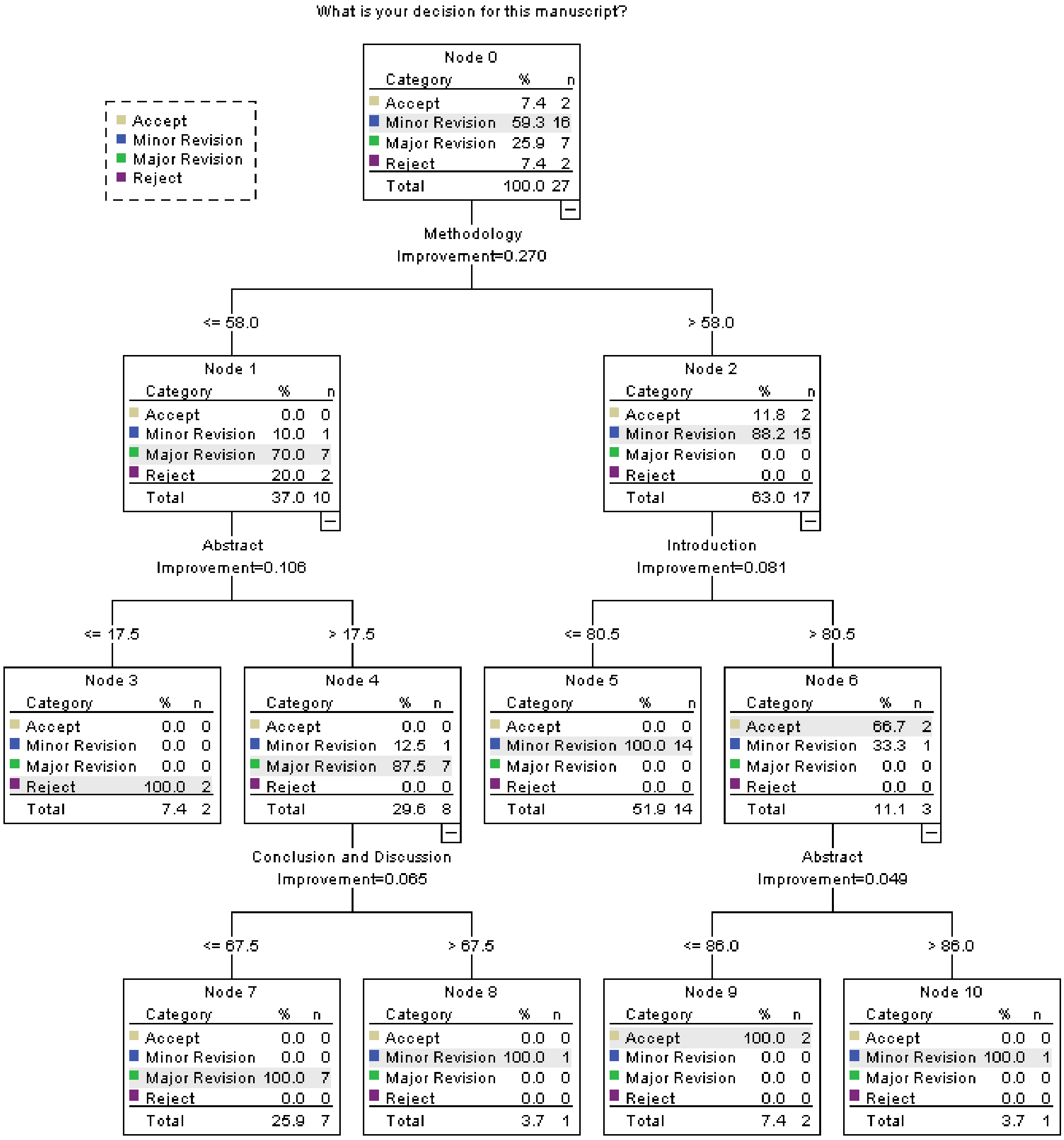


Figure 4-23 Decision Tree Using CRT for Schema Quality Attributes

As can be seen, out of these eight schema variables, four of them, Abstract, Introduction, Methodology and Conclusion and Discussion, can be seen here. The variable parts that are missing are mostly related to the ones that are associated with the references. This decision tree can be interpreted as follows:

1- The structure of the methodology is the main factor in predicting the decision for a manuscript. A reviewer who graded the selected manuscript of the study less than 58 (higher end of the Fair range) gave no chance to the manuscript for the direct acceptance and little chance of minor revision, with the paper ending up with major revision in this category.
2- Both “Introduction” and “Abstract” act as second level predictors. While the structure of the abstract is decisive in the rejection or revision of a paper, the structure of the introduction predicts the categorization of minor revisions.
3- Ultimately, there are “Conclusion and discussion” and “Abstract” that predict major revision and minor revision, or rate the manuscript for acceptance.

As can be seen, “Abstract” has prediction capability in two different paths. It is a high level predicting variable when the methodology is relatively weak, in order to reject a paper or allow the chance for revision. In another path, it is the final predictor to determine the acceptance of the selected manuscript.

Figure 4-24 shows the normalized importance of the schema variables. As can be seen, the range of diversity between the schema quality variables is more than the two previous scenarios of content quality attributes. However, for the schema variables related to the references, they are relatively similar to each other. On the other hand, adequate references, with more relevance to the literature review part rather than the reference part, has a lower level of importance in the normalized value. This diversity in the level of importance can be a hint on the better classification of the structural quality attributes in comparison with the content quality attributes. More explanation on certain insights from the range of importance of the quality attributes are provided in chapter 6.

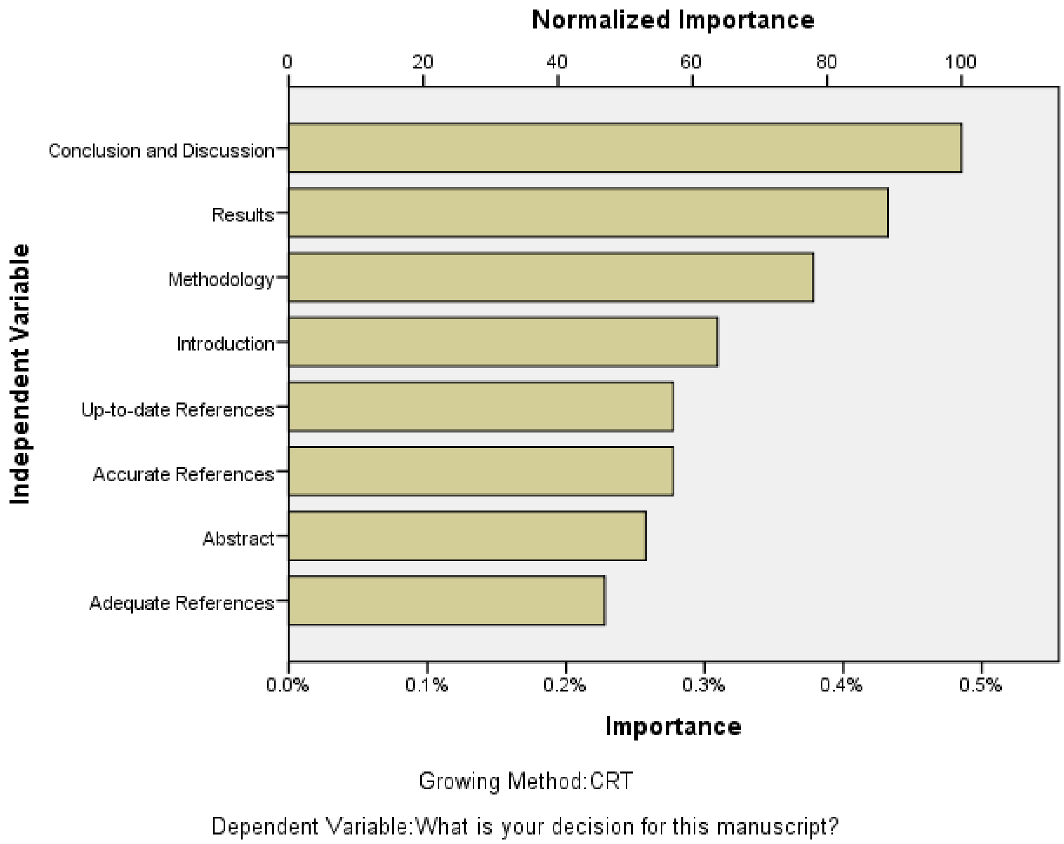


Figure 4-24 Normalized Importance of Schema Quality Variables

The performance of the decision tree on the schema quality attributes, similar to the content quality attributes, can be visualized through the gain percentage diagrams. Figure 4-25 shows the prediction performance for each of the target values using the schema quality attributes. As can be seen, the predictability performance of the schema attributes for a paper is relatively similar to the ones we obtained for the content quality attributes. This can show the structure of the mind-set of our study population that provides the same decision pattern under different quality attributes. In other words, the reviewer’s decision shows the same pattern for both the attributes that are associated with the schema and the attributes associated with the content. More explanation is provided in chapter 6.

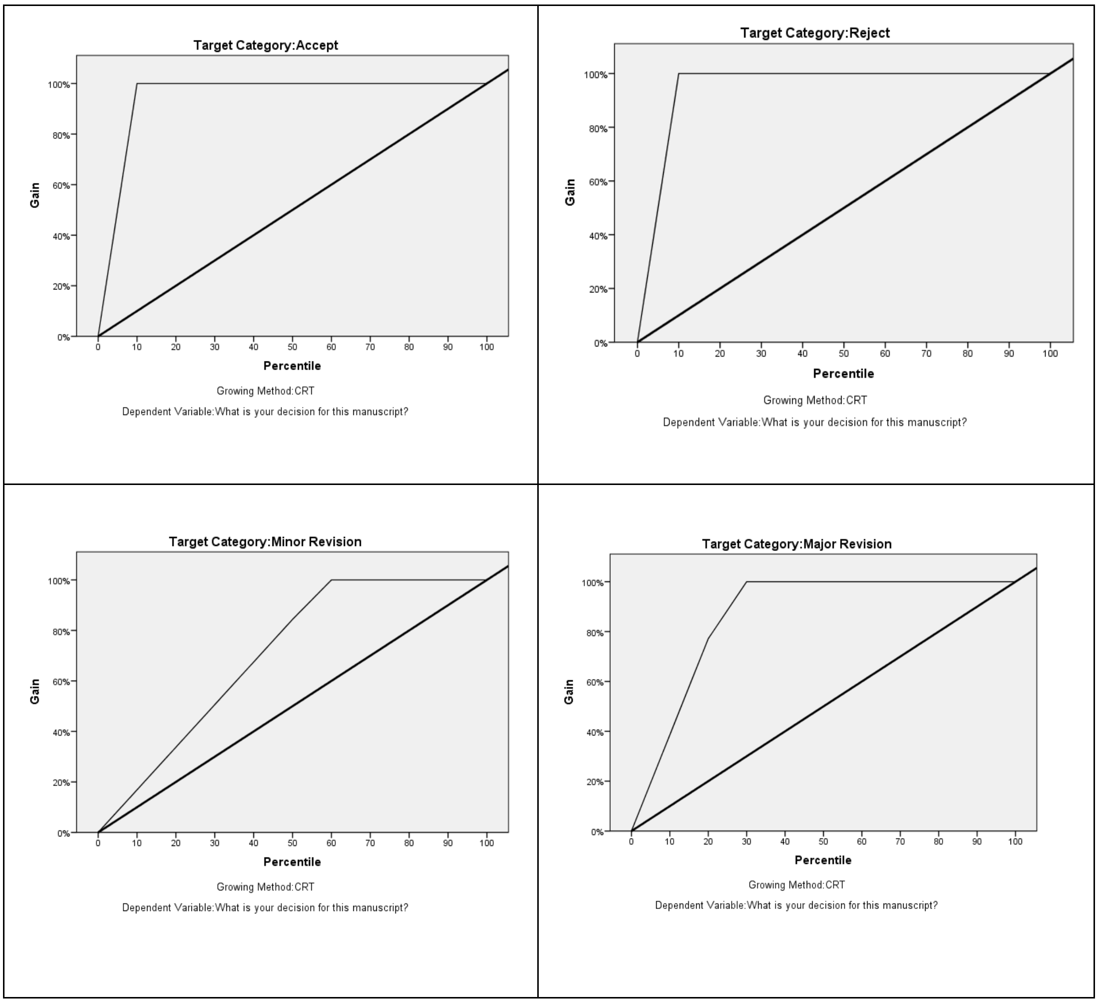


Figure 4-25 Gain Percentage for Decision Tree of Schema Quality Attributes

Finally, the decision paths derived from the schema attributes are given in Figure 4-26. This path summary suggests that the core structure of the manuscript, which is usually known as body of content, is the major predictor for most of the cases categorized as high or low quality papers, while the opening and closing parts of the manuscript can predict to what extent the manuscript may need revisions.

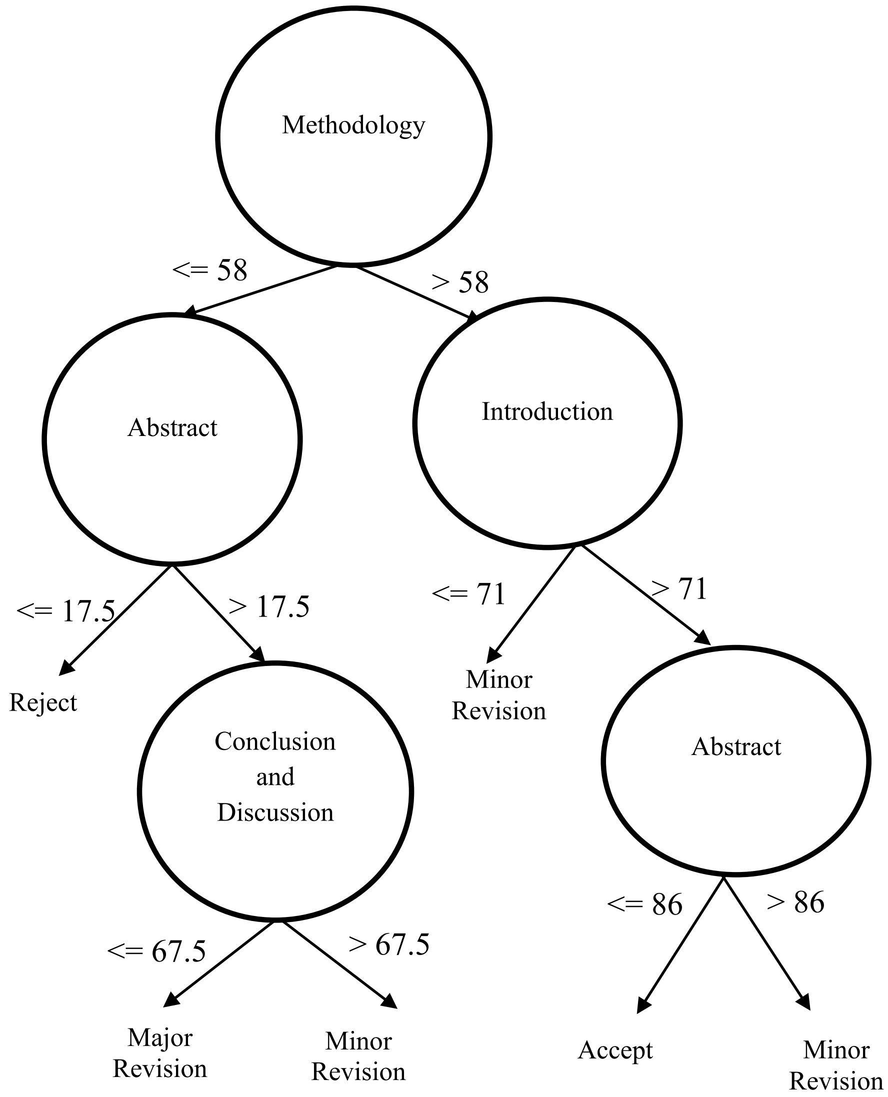


Figure 4-26 Decision Path for Schema Quality Attributes

4.6 Study on Content and Schema Preferences and Decision Patterns Using Focus Group Qualitative Approach

In the qualitative section of this study, we have conducted a focus group face to face interview with 5 experts in KM domain. Another KM paper was given to the participants to read before being given the assessment forms and decision tree cards. In order to have a more generalizable and reliable comparison, both the paper and journal were changed for this study. Similar to the quantitative approach, the assessment form comprised of quality attributes conducted with papers from Emerald and IGI-Global, plus the third section on the schema(format) attributes. (sample form available in appendix C).

4.6.1 Descriptive statistics for Quality Attributes in Qualitative Study

Based on the data collected form the participants. Table 4-31 shows the descriptive statistics obtained from the quality attributes of Emerald, IGI Global and Schema(format) from the assessment sheets given to the participants after they have read the selected paper.

As it can be seen from the descriptive data of the qualitative survey with 5 participants, for each scenario, we can compare the standard deviation of the quality attributes in order to better understand the diversity of each element among the five participants. We can summarise the diversities as follow:

A) Descriptive Analysis for the Emerald Scenario: In the Emerald Scenario, with total of six quality attributes, “Results” and “Implications for Research, practice and/or society” show the highest standard deviations, but their standard deviation is relatively close to each other. In the second level of standard deviation, “Relationship to Literature Review” and “Methodology” are relatively close to each other. Finally, we have "Originality" and "Quality of Communication and Language" with the lowest standard deviation.

Descriptive Statistics

		N	Minimum	Maximum	Mean	Std. Deviation
	Originality	5	50	70	60.00	7.071
	Relationship to Literature	5	55	80	66.00	10.840
	Methodology	5	50	80	63.00	10.954
Emerald	Result	5	40	80	60.00	14.577
	Implications for research, practice and/or society	5	40	80	63.00	14.832
	Quality of Communication & Language	5	70	80	76.00	5.477
	Popularity of the Subject	5	50	80	64.00	11.402
	Quality of research design	5	50	90	64.00	15.572
	Legitimacy of the conclusions	5	40	70	60.60	12.116
IGI	Appropriateness for the journal	5	75	100	83.00	9.747
	Adequacy of data analysis	5	40	90	57.00	19.235
	Practical/managerial significance	5	40	80	64.60	15.027
	Adequacy of literature review	5	50	80	64.00	15.166
	Contribution to the literature	5	50	70	60.60	7.197
	Clarity of presentation	5	65	80	71.00	5.477
	Abstract	5	50	85	66.00	12.942
	Introduction	5	50	75	64.00	9.618
	Methodology	5	50	90	61.60	17.686
	Results	5	40	80	63.00	14.832
	Conclusion and Discussion	5	40	70	59.00	13.416
Schema	Adequate References	5	63	80	74.60	7.797
	Accurate References	5	65	80	77.00	6.708
	Up-to-date References	5	65	80	72.00	5.701

Table 4-31 Descriptive Statistic Summary for Qualitative Survey

Looking back at the Figure 4-15 (page 98), it can be observed that “Results” is the starting point of the decision tree for Emerald scenario is our quantitative for the quantitative experiment. With the descriptive statistics here, we can see that the most significant diversity is grading based on the Emerald quality attributes also comes from the “Result” factor. Thus, the quality factor of “Results” can act as the most efficient classifier in the Emerald Scenario.

B) Descriptive Analysis the for IGI Scenario: For IGI scenarios, as table 4-31 shows, “Adequacy of data analysis” has the highest standard deviation. The next level of standard deviation is relatively close among “Adequacy of Literature”, “Quality of research design” and “Practical/Managerial Significance”. Finally, quality attributes of “Legitimacy of the conclusions”, “Popularity of the subject”, “Appropriateness for the Journal”, “Contribution to literature” and “Clarity of the Presentation” have the lower level of standard deviation in sequence. Again, by looking at the Figure 4-19 earlier in this chapter (page 105), it can be observed that “Adequacy of Data Analysis” is the starting point of IGI decision tree in quantitative analysis. In the qualitative approach, we can also find out that “Adequacy of Data Analysis” with the highest standard deviation which shows the diversity of this quality attributes among the participants in the focus group. Thus, “Adequacy of Data Analysis” can be considered as an efficient classifier in decision patterns.

C) Descriptive Analysis for the Schema (format): Finally, we have the quality attributes related to schema (format). We asked the participants in the interview to evaluate only the structure of the paper irrespective of the content. Looking at the standard deviations in this case, we can see that “Methodology” has the highest standard deviation among other quality attributes. Plus, the standard deviation among different quality attributes are spread widely, meaning the divergence of grading in some of the quality factors and a smoother convergence on some other factors. Looking back earlier in this chapter in Figure 4-23 (page 112), once again, it can be observed that methodology is the starting point of the decision tree in the quantitative approach and has the highest standard deviation in the qualitative study. Thus, "the schema(format) of the "Methodology" can be an efficient classifier in decision path. Plus, the better distribution of standard deviation in the qualitative study can suggest an evenly distributed tree for schema quality attributes in comparison with the quality attributes associated with the content.

4.6.2 Manual Decision Tree Building Using Plastic Cards in Qualitative Study

As mentioned earlier in Chapter 3, in the qualitative survey, instead of using the classifier software, we provided the participant with cards that were labelled with quality attributes used in the Emerald, IGI and Schema(format) scenarios. In each round, we gave each participant the set of cards related to each scenario. Thus, each participant should totally build three decision trees (Sample tree photo in appendix C). The participants chose their priorities by placing cards in vertical order (top-down). They could also arrange the cards horizontally for those quality attributes that they considered with same priority. Moreover, the participants were given the option not to use any quality attributes in case they don't find them significant in their decisions. Tables 4-32 to 4-34 show the quality attributes of each scenario for each participant and the position they took in decision trees built by the participants. For each scenario, the average priority is calculated and the lowest number (highest priority) will be marked with red circle.

As it can be seen, for the Emerald Scenario, "Methodology" has the highest rank among the participants for the focus group. Earlier, we showed that methodology is the starting point of the decision tree for the Emerald scenario in the quantitative and also showed that "Methodology" has the lowest standard deviation in the qualitative survey. Having the results from both qualitative and quantitative study, it can be concluded that "Methodology" plays a very significant role in decision path among all the other quality attributes represented in the Emerald scenario.

Emerald's Quality attributes	Originality	Relationship to literature review	Methodology	Result	Implications for Research, practice and/or Society	Quality of Communication & Language
Participant 1	2	1	1	1	3	2
Participant 2	1	1	2	3	1	-
Participant 3	1	6	2	3	4	5
Participant 4	1	2	2	2	2	3
Participant 5	2	-	1	-	-	-
Average Priority	7/4	10/5	8/5	9/4	10/4	10/4

Table 4-32 Priorities in Decision Tree for Emerald Scenario in Qualitative Study

IGI's Quality Attributes	Popularity of the Subject	Quality of research design	Legitimacy of the conclusions	Appropriateness for the journal	Adequacy of data analysis	Practical/manageria l significance	Adequacy of literature review	Contribution to the literature	Clarity of Presentation
Participant 1	4	1	3	2	1	2	2	4	3
Participant 2	-	2	3	-	2	1	1	1	-
Participant 3	4	1	7	3	2	6	8	9	5
Participant 4	3	1	1	2	1	2	2	1	2
Participant 5	-	2	-	-	1	-	-	-	-
Average Priority	11/3	7/5	14/4	7/3	7/5	11/4	13/4	15/4	10/3

Table 4-33 Priorities in Decision Tree for IGI Scenario in Qualitative Study

As it can be seen in Table 4-33, for the IGI scenario, the lowest average numbers (highest priority) are for two quality factors of "Quality of the research design" and "Adequacy of data analysis". Comparing the findings from the table with the decision trees and standard deviation discussed earlier in this chapter, it can be seen that "Adequacy of data analysis" is the starting point of the decision tree in the quantitative study and highest standard deviation in the qualitative study. At the same time, it has the top priority among the participants in the qualitative study. Thus, it can be concluded that the "Adequacy of data analysis" plays a significant role in the decision path.

Finally, Table 4-34 shows the priorities of participant in the qualitative study about their priorities about the structure and format (schema) of the paper for each section. Based on the average priority among the participants, it can be seen that the "Methodology" has the highest priority. Again, looking back at the quantitative survey on the schema (format) decision tree, we can observe that the decision tree for this scenario starts with "Methodology". The quality attribute of "Methodology" (as sown earlier) has also the highest standard deviation in the qualitative survey which makes a good candidate as an efficient classifier in decision path. Thus, it can be concluded that schema (format) of the methodology section in a manuscript plays a significant role in the decision about the manuscript.

Schema(format) Quality attributes	Abstract	Introduction	Methodology	Results	Conclusion and Discussion	Adequate References	Accurate References	Up-to-date Reference
Participant 1	4	4	1	2	4	2	2	3
Participant 2	-	1	2	3	3	-	-	-
Participant 3	4	5	1	2	3	6	8	7
Participant 4	1	1	2	1	2	3	3	3
Participant 5	-	-	1	2	-	-	-	-
Average Priority	9/3	11/4	7/5	10/5	12/4	11/3	13/3	13/3

Table 4-34 Priorities in Decision Tree for Schema Scenario in Qualitative Study

4.7 Summary of Knowledge Quality Decision Analysis in Peer Review

In chapter 4, we tried to analyse the dataset from different perspectives. In the first part of this chapter, we focused more on the descriptive analysis of the data in order to give a big picture about the background and preferences of the participants in the quantitative study, and furthermore, the relationship between different quality attributes and their distribution patterns were also shown in the form of cross-tabulation among different knowledge quality assessment criteria in peer-reviewing. In the second part of this chapter, decision tree analysis was introduced and discussed. Finally, based on both content and schema quality attributes analysed in the quantitative study, the decision tree for each scenario was generated and assessed.

In the last part of this chapter, we also introduced a second experiment in the form of qualitative study based on focus group interview and compared the findings of the qualitative experiment with the results of quantitative experiment discussed earlier in this chapter. Such comparison provided a better evidence in regards to the significance of knowledge quality attributes in each of the scenarios, providing a better reliability and generalizability of the decision trees derived from a set of quality attributes for each scenario.

In the next chapter, based on what has been explored about decision trees so far, we discuss the testing of the validity for decision patterns obtained from the decision tree in the quantitative study of this chapter. These validity tests are further explored to compare the decision paradigms when the sample size changes. Such analysis of change in the number of participants will open the door to explore between two types of peer-review, namely single-blind review and public open review. Such comparison will provide a better insight on the generalizability of peer-review decisions, which are explained in detail in chapter 6.

Chapter 5 : DECISION TREE VALIDATION APPROACH FOR PEER REVIEW COMPARISON

In chapter 4, we explored the descriptive part of the data for both qualitative and quantitative experiments. We also showed the decision paths using decision trees for different scenarios of knowledge quality assessment, for both content and schema in the quantitative experiment and endorsed the reliability of those decision trees by conducting a separate qualitative experiment in the form of focus group interview. While the performance of the decisions tree were discussed, the validity testing of the decision trees are also briefly introduced. One of the characteristics of the validation assessment test is to manipulate the sample size and structure in order to see if the same decision paths can be reached and if not, to what extent they might be different from the total size of the sample. Such characteristic will give a good opportunity, not only to assess the validation of decision path obtained from the decision trees for each scenarios, but also to provide good comparison assessment between peer-review types.

As mentioned in chapter two, there are various types of peer-review types, with the single blind reviews to be most popular one. On the other hand, as mentioned earlier, with the emergence of the internet and especially social media, there are more transparent approaches on larger scale decisions, which are known as public peer-review, that are assessed by a great number of experts within a scientific domain. One of the big differences of these recent approaches to peer-review is the size of the population giving decisions and feedback on a manuscript. This population size is much greater than in traditional peer-review processes that are limited to two to three people who give their opinions and shape the final destiny of the manuscript, to be published or rejected. Indeed, there are very few studies to evaluate the impact of decision size in assessing the quality of knowledge. Schultz's (2010) study is one of them that explored the convergence of decisions when the population of decision makers grow.

Bornmann & Daniel (2009) also studied different combinations of peer-review with three reviewers, but their focus was on different types of combination rather than manipulating (increase or decrease) the size of the population. Hargens (1988) and Hargens & Herting (1990) also discussed the fairness of decisions and the size of decisions. One of the very important factors that determines the final decision, based on the number of reviewers, as mentioned earlier in chapter 2, is the inter-rater reliability among reviewers (Bornmann, 2008; Marsh, Jayasinghe, & Bond, 2008; Rothwell & Martyn, 2000). This means how diversified the decisions from the reviewer are in regard to a manuscript.

In this chapter, based on the analysis we had in chapter 4, we explore the different aspects of such inter-rater reliability using validation and performance measures from decision trees.

5.1 Decision Scaling Using Decision Tree Validation Methods

In SPSS, which has been used for analysis in this study, there are mainly two types of validation techniques to assess the performance of decision trees, as briefly discussed in Chapter 4. These two methods are 1- *Crossvalidation* and 2-*Split-sample Validation*. We explain each method first, and then check each validation method for the three scenarios discussed in chapter 4.

5.1.1 Knowledge Quality Decision Clustering Using the Crossvalidation Method

In SPSS, crossvalidation is the validation mechanism that breaks the dataset into a number of folds or segments. Figure 5-1 show the setup for the crossvalidation method in SPSS.

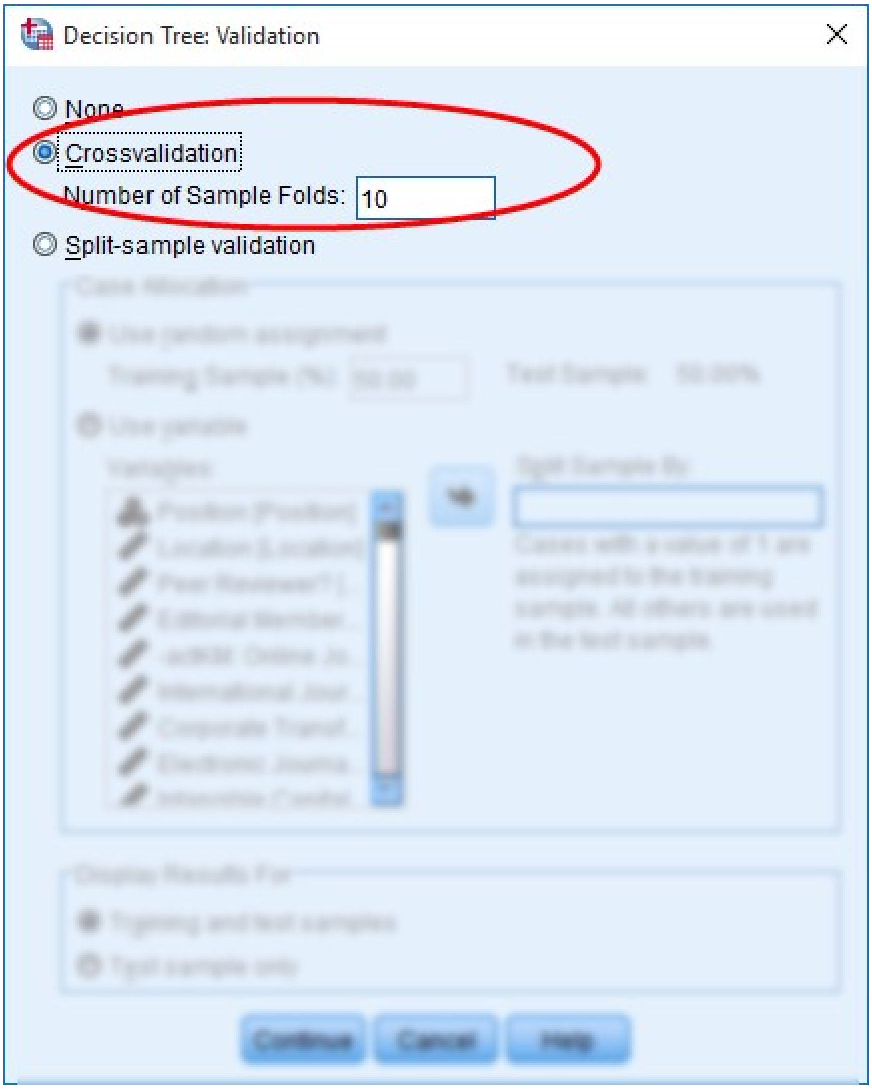


Figure 5-1 Crossvalidation Setup for Decisions Trees in SPSS

As can be seen, crossvalidation has only one variable which is named as “Fold”. Fold in fact is the number of subsamples. The working mechanism for crossvalidation is as follows:

1- The program receives the number of folds from the system. Based on the number of folds, which ranges between 2 to a maximum of 25 (in SPSS), the program generates decision trees by excluding the cases (reviewers in this study) of the other folds. Apparently, as the total dataset size is not growing (27 participants in this study), the greater the number of folds, the smaller the number of cases that will be excluded in validation.
2- Based on the number of folds, the software will generate a series of decision trees where each decision tree is based on all cases in the dataset, excluding the cases assigned to its subsample. Thus, the first tree is based on all cases except the one assigned to subsample one, the second tree is based on all cases except those assigned to subsample two and so on. This create the opportunity in each decision tree to lose the data from part of the sample size and construct itself based on the rest of the dataset information. In the format of this study, this means that the decision tree in each round is losing decision information from a subsample of the reviewers. For this study with the dataset of 27 participants, having 2 folds means, each decision tree loses the decision information of roughly 13 or 14 (27/2 ~13 or 14) participants in constructing decision trees in each round.
3- The creation of decision tree continues up to the last fold, meaning, that the number of decision trees produced matches the number of folds. However, the final output from crossvalidation will be a single tree which carries the average risk values of all decision trees produced for each fold (subsample).

The decision summary received as a decision tree with the average risk of all trees gives an insight on how the decision path may vary by removing a number of reviewers from the decision (assuming the final decision is based on the counts of the most frequent decision). The crossvalidation gives an opportunity to see the variation level in decision path of the same manuscript submitted to a completely different set of reviewers (for number of folds when set to 2) and the variation decision when some reviewers are excluded (for number of fold of 3 and above).

5.1.2 Knowledge Quality Decision Training and Testing Using Split-Sample Method

In the split-sample validation method, instead of clustering the dataset into subsamples, we separate a portion of the dataset based on a percentage of the whole dataset and categorize it into being either the “Test” or the “Training” subset. Figure 5-2 shows the setup for the split-sample method in SPSS. As can be seen, the percentage of the training sample size can be set and the percentage of test sample will change accordingly, with consideration the maximum total of 100% (e.g a 20% training sample will have a 80% testing sample, 30% training will have 70% test etc.). One of the characteristics of the split-sample model is its various testing characteristics based on the percentage of the training sample. A training sample with a small percentage can test the generalizability of the decision tree. A sample test of equal percentages may provide an insight in the comparison and reliability of similar groups (in size) and a large training percentage may provide some statistical capabilities into sampling decisions for a large scale dataset. It should be noted that for small samples (like in our study), very small training samples need caution, as too few cases for training may fail to result in any tree to be formed.

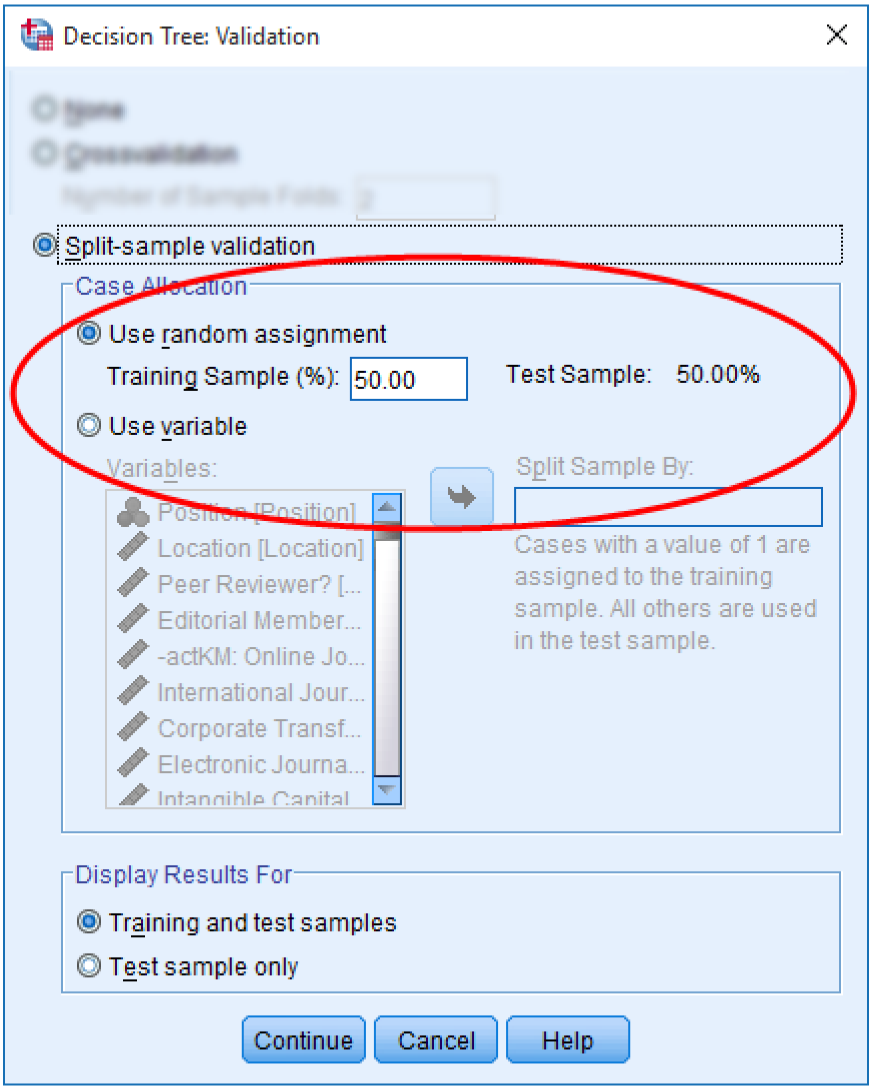


Figure 5-2 Split-sample Validation Setup for Decision Trees in SPSS

The working mechanism of the Split-sample is as follows:

1- The program receives the percentage of the training sample from the whole dataset and adjusts the testing sample percentage accordingly. A small percentage training sample results in a larger percentage of testing sample and vice versa.

2- On the basis of the training sample, the program tries to find a decision tree with the lowest risk (based on the given growth method) and applies the structure of the decision tree obtained from the trained subset to the testing sample.

3- Finally, it assesses the risk based on the coverage of the testing sample on the structure of the decision tree derived from the training sample.

One of the options to be set in the split-sample method is to either choose the training sample from a fixed set of cases in a dataset or let the program randomly assign the cases in each run. For this study, each case (reviewer) is treated the same on the decision so randomized assignment will help in assessing the decision holistically rather than on a case basis.

With this introduction to the validation mechanism, we explore each of these validation methods into all scenarios given in chapter 4. In order to minimise the bias with the discussion and analysis of chapter two, we do all the validation testing with the similar growth method used in the scenarios in chapter 4 which, for all cases, was chosen to be CRT.

5.2 Decision Tree Validation Analysis for Content Quality Criteria Adopted from Emerald

In the first scenario of chapter 4, we constructed the decision tree based on the six quality content attributes adopted from the publisher Emerald. As it was seen in Figure 4-15, out of the 6 variables, the three variables of Results, Originality and Relationship to Literature Review could classify the dataset 100% with the use of CRT growth method. In this section, we are going to apply both Crossvalidation and Split-sample validation to assess the structure of the decision tree constructed for the whole dataset obtained from the survey.

5.2.1 Crossvalidation of Decision Tree Constructed from Content Quality Criteria Adopted from Emerald

In order to properly assess the effect of decision size on the decision tree, we apply different levels of clustering by changing the number of folds between 2 to 5. In such a fold setting we roughly eliminate 5 to 14 (~27/5 to ~27/2) reviewers in constructing the decision trees. For each fold, as the dataset case assignment is random for each run, we run the program 10 times for each fold setting to get a better average for the risk level from ten iterations, rather than merely relying on the risk assessment for one run. Table 5-1 shows the risk estimate for 10 iterations of random clustering for 2 folds.

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.185	0.075	Results	Originality	Originality	Relationship to Literature
2	0.259	0.084	Results	Originality	Originality	Relationship to Literature
3	0.296	0.088	Results	Originality	Originality	Relationship to Literature
4	0.333	0.091	Results	Originality	Originality	Relationship to Literature
5	0.333	0.091	Results	Originality	Originality	Relationship to Literature
6	0.259	0.084	Results	Originality	Originality	Relationship to Literature
7	0.259	0.084	Results	Originality	Originality	Relationship to Literature
8	0.407	0.095	Results	Originality	Originality	Relationship to Literature
9	0.148	0.068	Results	Originality	Originality	Relationship to Literature
10	0.222	0.08	Results	Originality	Originality	Relationship to Literature
Average	0.27	0.084				

Table 5-1 Crossvalidation of Emerald's Content Quality Attributes with 2 Folds

As the table 5-1 shows, for 2 folds (meaning separating the sample size into two subsets), by randomly choosing the cases (reviewers for this study) into these two groups, we expect that, on average (after 10 iterations), we get a risk average of 0.27 with a standard deviation of 0.076. Plus, the predictor priorities are also shown and are similar across the validation for the 10 iterations. These predictor priorities are also similar to these tested in the original dataset in the previous chapter and shown in Figure 4-15. From Table 5-1, it can be interpreted that, by classifying the decision tree in two separate folds (each fold will have around 13 or 14 cases from total of 27), we can expect that the decision path shown in Figure 4-18 in chapter 4, can predict, on average, around 73% of the decisions (1- average of risk estimate). We continue the validation test for the 3, 4 and 5 folds as shown in table 5-2, table 5-3 and table 5-4 respectively.

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.296	0.088	Results	Originality	Originality	Relationship to Literature
2	0.37	0.093	Results	Originality	Originality	Relationship to Literature
3	0.37	0.093	Results	Originality	Originality	Relationship to Literature
4	0.333	0.091	Results	Originality	Originality	Relationship to Literature
5	0.259	0.084	Results	Originality	Originality	Relationship to Literature
6	0.296	0.088	Results	Originality	Originality	Relationship to Literature
7	0.259	0.084	Results	Originality	Originality	Relationship to Literature
8	0.259	0.084	Results	Originality	Originality	Relationship to Literature
9	0.148	0.068	Results	Originality	Originality	Relationship to Literature
10	0.37	0.093	Results	Originality	Originality	Relationship to Literature
Average	0.296	0.086				

Table 5-2 Crossvalidation of Emerald’s Content Quality Attributes with 3 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.37	0.093	Results	Originality	Originality	Relationship to Literature
2	0.222	0.08	Results	Originality	Originality	Relationship to Literature
3	0.333	0.091	Results	Originality	Originality	Relationship to Literature
4	0.259	0.084	Results	Originality	Originality	Relationship to Literature
5	0.259	0.084	Results	Originality	Originality	Relationship to Literature
6	0.37	0.093	Results	Originality	Originality	Relationship to Literature
7	0.259	0.084	Results	Originality	Originality	Relationship to Literature
8	0.296	0.088	Results	Originality	Originality	Relationship to Literature
9	0.296	0.088	Results	Originality	Originality	Relationship to Literature
10	0.259	0.084	Results	Originality	Originality	Relationship to Literature
Average	0.292	0.086				

Table 5-3 Crossvalidation of Emerald's Content Quality Attributes with 4 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.222	0.08	Results	Originality	Originality	Relationship to Literature
2	0.296	0.088	Results	Originality	Originality	Relationship to Literature
3	0.259	0.084	Results	Originality	Originality	Relationship to Literature
4	0.259	0.084	Results	Originality	Originality	Relationship to Literature
5	0.296	0.088	Results	Originality	Originality	Relationship to Literature
6	0.296	0.088	Results	Originality	Originality	Relationship to Literature
7	0.333	0.091	Results	Originality	Originality	Relationship to Literature
8	0.296	0.088	Results	Originality	Originality	Relationship to Literature
9	0.296	0.088	Results	Originality	Originality	Relationship to Literature
10	0.333	0.091	Results	Originality	Originality	Relationship to Literature
Average	0.288	0.087				

Table 5-4 Crossvalidation of Emerald's Content Quality Attributes with 5 Folds

As it can be seen from Table 5-2 to Table 5-4, while the predictor priorities remain similar, the average estimated risk changes. These average risks are 0.296, 0.292 and 0.288 for 3, 4 and 5 folds respectively. These numbers show, on the average of 10 iterations with risk assessments, it is expected that with 2,3,4 and 5 folds, the decision tree originating from these content quality attributes can predict between 71% to 73% of the cases. Figure 5-3 shows these prediction variations. Such a small variation of the predictability percentage from different sets of fold numbers suggests that the decision path shows little sensitivity with reducing or expanding the number of cases (reviewers) for this study. A detailed explanation of the analysis is given for all scenarios in chapter 6.

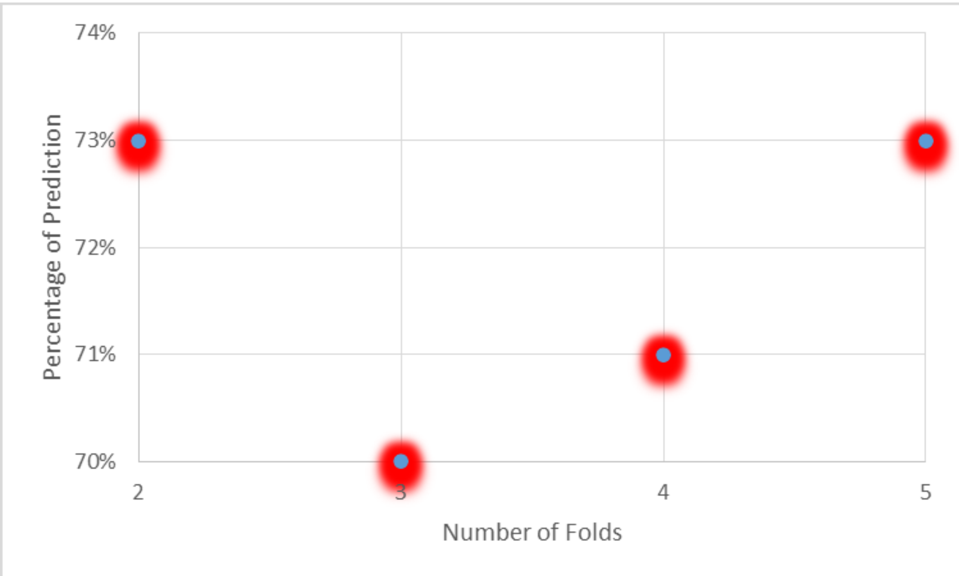


Figure 5-3 Crossvalidation Prediction Rate for Different Folds of Emerald's content Quality Attributes

5.2.2 Split-sample Validation of Decision Tree Constructed from Content Quality Criteria Adopted from Emerald

As we explored the crossvalidation method in the previous part, in this section, we apply the second validation method, the split-sample method. As mentioned earlier, in the split-sample method, instead of folds, we have the percentages for testing and training the sample. In this study, we set the training set percentage to 20%, 50% and 80%, with corresponding test percentages of 80%, 50% and 20%. It should be noted that, as long as the dataset is relatively small, using a very low percentage for training may not provide a desirable decision tree solution. We also run the test at each training percentage 10 times in order to get a better insight on risk and pattern changes. Table 5-5 to Table 5-7 shows the testing results for training sets of 20%, 50% and 80% in 10 runs. It also shows the predictor's priority for the each iteration based on the decision tree output that is obtained for each iteration from the training set.

As it can be seen from Table 5-5, the average estimated risk is 0.354. This means, on average, by having 20% of our dataset for training, we could predict roughly 65% of the decisions of the testing set. The average predicted error is 0.161 (16%). In table 5-5, for 50% training, on average, we had 0.293 estimated risk with an average standard error of 0.121 which gives around 71% predictability in 10 iterations for the testing set, with 12% of average standard error. Finally, in Table 5-7, for the 80% training set, we achieved an average estimated risk of 0.326, with an average standard error of 0.159. This can be interpreted that 68% of the decisions of 20% of the population can be predicted by 80% in the training set and with roughly around 16% standard error.

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority			
				1	2	3	4
1	Training	0	0	Originality	-	-	-
	Testing	0.35	0.107				
2	Training	0	0	Results	-	-	-
	Testing	0.63	0.061				
3	Training	0	0	Results	-	-	-
	Testing	0.63	0.061				
4	Training	0	0	Results	-	-	-
	Testing	0.118	0.078				
5	Training	0	0	Originality	-	-	-
	Testing	0.118	0.078				
6	Training	0	0	Originality	-	-	-
	Testing	0.286	0.171				
7	Training	0	0	Methodology	Originality	Originality	-
	Testing	0.25	0.097				
8	Training	0	0	Implications for Research & Practise	Originality	-	-
	Testing	0.786	0.064				
9	Training	0	0	Originality	-	-	-
	Testing	0.167	0.88				
10	Training	0	0	Originality	-	-	-
	Testing	0.211	0.094				
Average	Training	0	0				
	Testing	0.354	0.1691				

Table 5-5 Split-sample Validation of Emerald's Content Quality Attributes with 20% Training

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority			
				1	2	3	4
1	Training	0	0	Relationship to literature review	Originality	Originality	-
	Testing	0.462	0.138				
2	Training	0	0	Results	Originality	Originality	-
	Testing	0.083	0.08				
3	Training	0	0	Results	Originality Methodology	-	-
	Testing	0.273	0.134				
4	Training	0	0	Results	Originality	Originality	-
	Testing	0.444	0.166				
5	Training	0	0	Originality	Originality	Originality	-
	Testing	0.429	0.132				
6	Training	0	0	Relationship to literature review	Originality	Originality	-
	Testing	0.333	0.136				
7	Training	0	0	Methodology	Originality	Originality	-
	Testing	0.267	0.114				
8	Training	0	0	Methodology	-	-	-
	Testing	0.091	0.087				
9	Training	0	0	Originality	Originality	Originality	-
	Testing	0.267	0.114				
10	Training	0	0	Results	Originality	Originality	-
	Testing	0.286	0.121				
Average	Training	0	0				
	Testing	0.293	0.121				

Table 5-6 Split-sample Validation of Emerald's Content Quality Attributes with 50% Training

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority			
				1	2	3	4
1	Training	0	0	Results	Originality	Originality	Relationship to literature review
	Testing	0	0				
2	Training	0	0	Results	Originality Methodology	Originality	Relationship to literature review Originality
	Testing	0.5	0.177				
3	Training	0	0	Relationship to literature review	Originality	Originality	-
	Testing	0.4	0.155				
4	Training	0	0	Originality	Originality Methodology	Originality	-
	Testing	0.5	0.177				
5	Training	0	0	Results	Originality	Originality	-
	Testing	0.167	0.152				
6	Training	0	0	Methodology	Originality	Originality	-
	Testing	0.25	0.153				
7	Training	0	0	Results	Originality	-	-
	Testing	0.5	0.177				
8	Training	0	0	Results	Originality	Originality	Relationship to literature review
	Testing	0.2	0.179				
9	Training	0	0	Results	Originality	Originality	-
	Testing	0.5	0.204				
10	Training	0	0	Results	Originality Methodology	Originality	Relationship to literature review
	Testing	0.25	0.217				
Average	Training	0	0				
	Testing	0.326	0.159				

Table 5-7 Split-sample Validation of Emerald's Content Quality Attributes with 80% Training

Comparing the data achieved from ten iteration of these three percentages for the training set, Figure 5-4 shows the variation of the percentage predictability.

Figure 5-4 Split-Sample Validation Ratio for Emerald's content Quality Attributes

From both validation tests on the content quality attributes adopted from Emerald, it can be seen, while the sample size has an effect on the percentage of predictability, the changes in the rate of prediction is not very significant. Nevertheless, there are changes in the structure of the decision tree based on the size of the sample that might give high or low levels of prediction in different circumstances. The impact of rapid changes will also increase when the sample size is small and the majority of the outliers fall in the random training set.

5.3 Decision Tree Validation Analysis for Content Quality Criteria Adopted from IGI

We discussed validation analysis of the content quality criteria adopted from Emerald in the previous section. In this part, we apply the same type of validation, but, we now choose the 9 variables that have been adopted from IGI-global. Similar to the previous section, we choose the CRT growth method as the best predictor for this scenario too.

5.3.1 Crossvalidation of Decision Tree Constructed from Content Quality Criteria adopted from IGI

Similar to the tests in the previous section, we do the crossvalidation for different numbers of folds, from 2 to 5 on the 9 variables adopted from IGI. For each fold, we also run the crossvalidation test 10 times to get a more accurate average for the output. Table 5-8 to Table 5-11 show the estimated risk and predictor priorities achieved in each iteration. As it can be seen from these figures, the prediction rate for 2, 3, 4 and 5 folds are roughly 73%, 78%, 76% and 75%. In fact, similar to Emerald scenario in the previous section, while there is a variation in the prediction level, such variation is limited to relatively a small percentage. Comparing the result with the previous scenario, having 5 main predictors for IGI (in comparison with 4 in Emerald), the average prediction rate, based on IGI content quality attributes, is slightly higher than those obtained from the ones adopted from Emerald. Nevertheless, the prediction rate for both scenarios of cross-tabulations stayed between 70% and 78%, on the average, obtained after 10 iterations.

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.296	0.088	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
2	0.148	0.068	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
3	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
4	0.37	0.093	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
5	0.296	0.088	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
6	0.407	0.095	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
7	0.333	0.091	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
8	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
9	0.148	0.068	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
10	0.222	0.08	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
Average	0.266	0.083				

Table 5-8 Crossvalidation of IGI's Content Quality Attributes with 2 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
2	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
3	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
4	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
5	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
6	0.222	0.08	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
7	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
8	0.296	0.088	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
9	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
10	0.222	0.08	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
Average	0.218	0.0791				

Table 5-9 Crossvalidation of IGI's Content Quality Attributes with 3 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.333	0.091	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
2	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
3	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
4	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
5	0.296	0.088	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
6	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
7	0.222	0.08	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
8	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
9	0.111	0.06	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
10	0.333	0.091	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
Average	0.244	0.0812				

Table 5-10 Crossvalidation of IGI's Content Quality Attributes with 4 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.296	0.088	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
2	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
3	0.222	0.08	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
4	0.333	0.091	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
5	0.296	0.088	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
6	0.222	0.08	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
7	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
8	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
9	0.259	0.084	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
10	0.185	0.075	Adequacy of data Analysis	Popularity of the Subject Quality of Research Design	Popularity of the subject	Adequacy of data Analysis
Average	0.251	0.0829				

Table 5-11 Crossvalidation of IGI's Content Quality Attributes with 5 Folds

Figure 5-5 shows the variation range with different number of fold for IGI scenario.

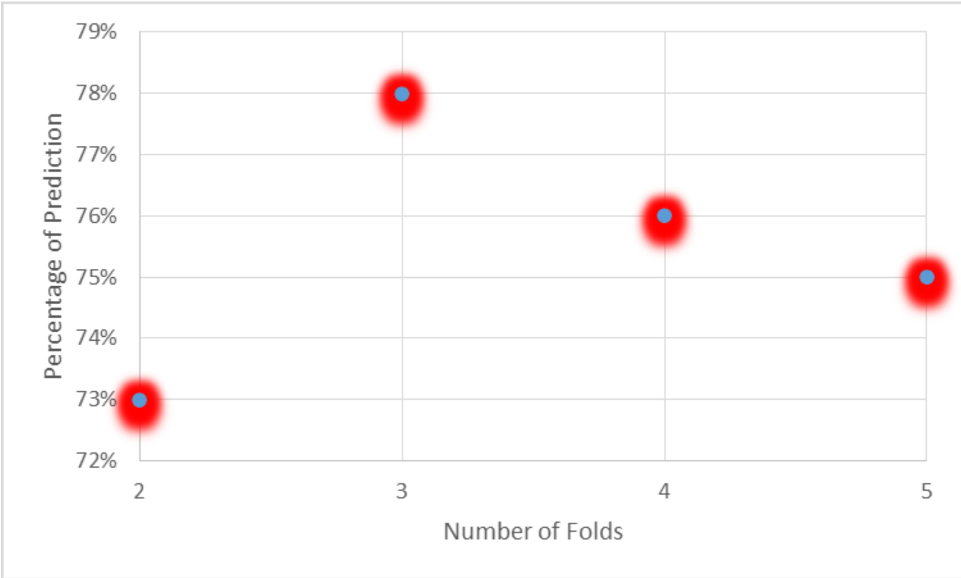


Figure 5-5 Crossvalidation Prediction Rate for Different Folds of IGI's content Quality Attributes.

5.3.2 Split-sample Validation of the Decision Tree Constructed from Content Quality Criteria adopted from IGI

In this part, similar to the split-sample validation test we did for Emerald, we apply the same settings for the 9 variables adopted from IGI. The training percentages of 20%, 50% and 80% were applied and the predictors are listed based on their priorities in each training set decision tree. Table 5-12 to Table 5-15 show the average risk of both the testing and training sets and the corresponding standard error, followed by the decision tree attribute priorities.

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority			
				1	2	3	4
1	Training	0	0	Adequacy of Literature Review	-	-	-
	Testing	0.118	0.078				
2	Training	0	0	Appropriateness of the journal	Popularity of the subject	-	-
	Testing	0.5	0.118				
3	Training	0	0	Adequacy of Literature Review	Popularity of the subject	-	-
	Testing	0.571	0.132				
4	Training	0	0	Appropriateness of the journal	Popularity of the subject	-	-
	Testing	0.25	0.097				
5	Training	0	0	Popularity of the subject	-	-	-
	Testing	0.4	0.126				
6	Training	0	0	Adequacy of Literature Review	-	-	-
	Testing	0.4	0.219				
7	Training	0	0	Popularity of the subject	Popularity of the subject	-	-
	Testing	0.316	0.107				
8	Training	0	0	Popularity of the subject	-	-	-
	Testing	0.35	0.107				
9	Training	0	0	Popularity of the subject	Popularity of the subject	-	-
	Testing	0.471	0.121				
10	Training	0	0	Quality of Research Design	-	-	-
	Testing	0.067	0.064				
Average	Training	0	0				
	Testing	0.344	0.116				

Table 5-12 Split-sample Validation of IGI's Content Quality Attributes with 20% Training

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority 1	Predictor Priority 2	Predictor Priority 3	Predictor Priority 4
1	Training	0	0	Adequacy of data analysis	Popularity of the subject Contributions to Literature	Popularity of the subject	-
	Testing	0.091	0.087				
2	Training	0	0	Adequacy of data analysis	Popularity of the subject Appropriateness of the journal	Quality of research design	-
	Testing	0.571	0.132				
3	Training	0	0	Adequacy of data analysis	Popularity of the subject	Quality of research design	Popularity of the subject
	Testing	0.154	0.1				
4	Training	0	0	Adequacy of data analysis	Popularity of the subject	-	-
	Testing	0	0				
5	Training	0	0	Adequacy of data analysis	Popularity of the subject	-	-
	Testing	0.364	0.145				
6	Training	0	0	Quality of research design	Popularity of the subject Appropriateness of the journal	Appropriateness of the journal	-
	Testing	0.5	0.134				
7	Training	0	0	Quality of research design	Popularity of the subject	-	-
	Testing	0.364	0.145				
8	Training	0	0	Quality of research design	Popularity of the subject	Popularity of the subject	-
	Testing	0.357	0.128				
9	Training	0	0	Adequacy of data analysis	Contributions to Literature	Popularity of the subject	-
	Testing	0.077	0.074				
10	Training	0	0	Appropriateness of the journal	Popularity of the subject Quality of research design	Popularity of the subject	-
	Testing	0.333	0.122				
Average	Training	0	0				
	Testing	0.28	0.106				

Table 5-13 Split-sample Validation of IGI's Content Quality Attributes with 50% Training

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority 1	Predictor Priority 2	Predictor Priority 3	Predictor Priority 4
1	Training	0	0	Adequacy of data analysis	Popularity of the subject Quality of research design	-	-
	Testing	0.5	0.25				
2	Training	0	0	Adequacy of data analysis	Popularity of the subject	Popularity of the subject	-
	Testing	0.167	0.152				
3	Training	0	0	Quality of research design	Popularity of the subject Quality of research design	Quality of research design	-
	Testing	0.333	0.192				
4	Training	0	0	Adequacy of data analysis	Popularity of the subject	Appropriateness of the journal	Appropriateness of the journal
	Testing	0.286	0.171				
5	Training	0	0	Adequacy of data analysis	Popularity of the subject Quality of research design	Contribution to literature	Appropriateness of the journal
	Testing	0.2	0.179				
6	Training	0	0	Adequacy of data analysis	Popularity of the subject Quality of research design	Popularity of the subject	Adequacy of data analysis
	Testing	0	0				
7	Training	0	0	Adequacy of data analysis	Popularity of the subject Quality of research design	Quality of research design	-
	Testing	0.2	0.179				
8	Training	0	0	Quality of research design	Popularity of the subject	Adequacy of data analysis	Quality of research design Appropriateness of the journal (**5th priority**)
	Testing	0.5	0.25				
9	Training	0	0	Adequacy of data analysis	Popularity of the subject Quality of research design	Quality of research design	-
	Testing	0.25	0.217				
10	Training	0	0	Adequacy of data analysis	Popularity of the subject	Adequacy of data analysis	Quality of research design
	Testing	0.667	0.272				
Average	Training	0	0				
	Testing	0.31	0.191				

Table 5-14 Split-sample Validation of IGI's Content Quality Attributes with 80% Training

As it can be seen from table 5-12 to 5-14, the average estimated risks for training of 20%, 50% and 80% training with 10 iterations are 0.344, 0.28 and 0.31. These risk estimates are equal to predictability levels of 66%, 72% and 69%. Figure 5-6 shows the predictability rate based on the level of training for content quality attributes adopted from IGI.

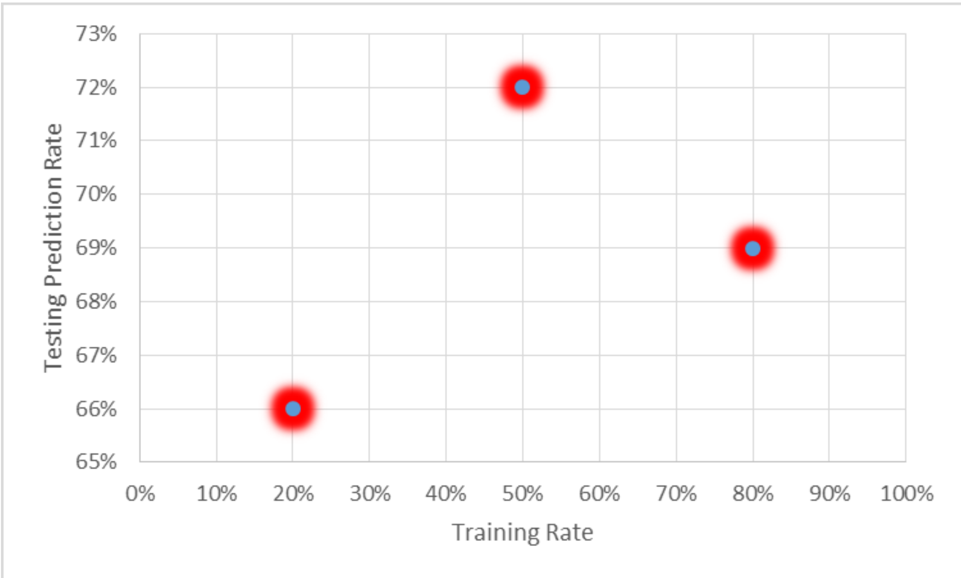


Figure 5-6 Split-Sample Validation Ratio for IGI's content Quality Attributes

By comparing the validation analysis of the content quality factors that were adopted from both Emerald and IGI, it can be see that, although there is a variation in predictability based on the training size, nevertheless, such variation may not be very significant if we try to compare a series of iterations rather than comparing on a case to case basis. In fact, having more content quality variables (in the case of IGI), or changes in sample size, made some changes to the level of predictability of the decision path, but these changes ranged not more than 10% overall in the best and the worst scenarios. More explanation on the result of validation for the content quality is given in chapter 6.

5.4 Decision Tree Validation Analysis for Schema Quality Criteria

In the previous section, we analysed the validation trees for content quality attributes obtained from both IGI and Emerald. In chapter 4, we explored the decision trees for the schema quality attributes too. In this section, similar to the validation for content quality attributes, we do both crossvalidation and split-sample validation tests with similar setups to the validation of the content quality attributes. All the validation of this section is be based on the 8 schema quality attributes listed in Table 4-14. Moreover, as the CRT growth was the best performing method on making the decision tree for the schema quality attributes, as discussed in chapter 4, we will adopt the same growth method for validation of the schema variables in this section too.

5.4.1 Crossvalidation of Decision Tree constructed from Schema Quality

For the crossvalidation of the schema attributes, similar to quality content variables, we will make 2, 3,4 and 5 folds from the dataset and run it 10 times to get a better prediction average for each fold. Table 5-15 to Table 5-18 show the estimated risk achieved in each iteration for each number of folds, based on the schema quality attributes. These estimated risk values for 2, 3, 4 and 5 folds are 0.262, 0.251, 0.214 and 0.218 respectively, approximately reflecting prediction levels of 74%, 75%, 79% and 78%.

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
2	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
3	0.296	0.088	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
4	0.481	0.096	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
5	0.444	0.096	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
6	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
7	0.148	0.068	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
8	0.296	0.088	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
9	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
10	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
Average	0.262	0.081				

Table 5-15 Crossvalidation of Schema Quality Attributes with 2 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
2	0.333	0.091	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
3	0.333	0.091	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
4	0.333	0.091	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
5	0.296	0.088	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
6	0.259	0.084	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
7	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
8	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
9	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
10	0.148	0.068	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
Average	0.251	0.0823				

Table 5-16 Crossvalidation of Schema Quality Attributes with 3 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.296	0.088	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
2	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
3	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
4	0.259	0.084	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
5	0.259	0.084	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
6	0.148	0.068	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
7	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
8	0.259	0.084	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
9	0.148	0.068	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
10	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
Average	0.214	0.070				

Table 5-17 Crossvalidation of Schema Quality Attributes with 4 Folds

Iteration #	Risk Estimate	Standard Error	Predictor Priority			
			1	2	3	4
1	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
2	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
3	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
4	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
5	0.222	0.08	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
6	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
7	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
8	0.185	0.075	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
9	0.259	0.084	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
10	0.296	0.088	Methodology	Abstract Introduction	Conclusion & Discussion Abstract	-
Average	0.218	0.079				

Table 5-18 Crossvalidation of Schema Quality Attributes with 5 Folds

The prediction values obtained from the crossvalidation tables for schema quality attributes are mapped in Figure 5-7. The variation of prediction is between 74% and 79%. By comparing Figure 5-7 with the results obtained from the content quality attributes from both IGI and Emerald, It can be seen that the variation is relatively narrow inside each category type (approximately 5% to 6%). However, when comparing the minimum percentage of predictability in the schema quality attributes with those obtained from the content quality attributes, the minimum prediction level of the schema attributes is higher than the maximum we obtained from Emerald for 10 iterations. On the other hand, the prediction levels in the schema quality attributes and the ones obtained from the IGI content quality attributes are relatively within the same range. Further discussion on the comparison for the prediction rates of the content and schema is given in Chapter 6.

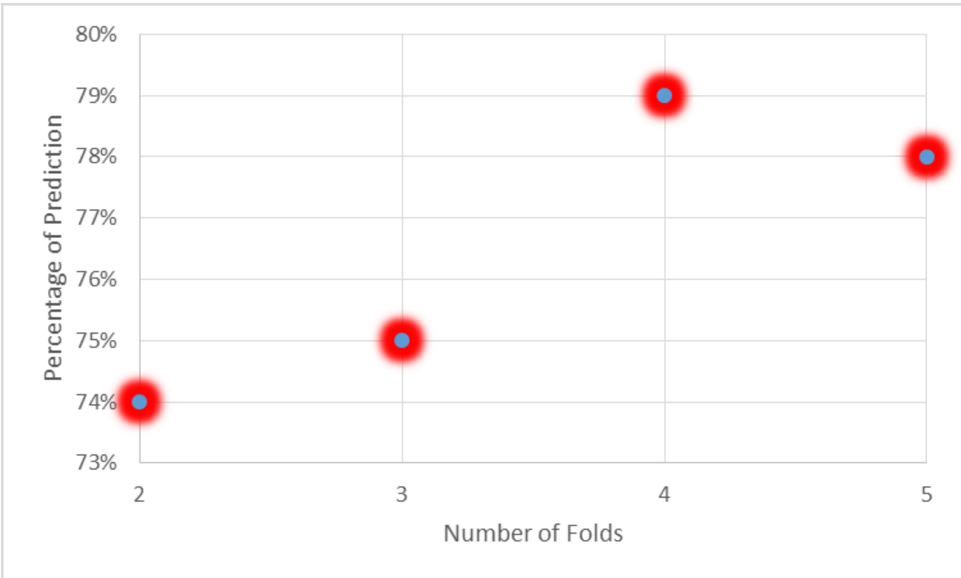


Figure 5-7 Crossvalidation Prediction Rate for Different Folds of Schema Quality Attributes.

5.4.2 Split-sample Validation of Decision Tree constructed from Schema Quality Attributes

The final validation testing is the split-sample validation for the schema quality attributes. Once again, similar to that of the content quality attributes, the validation takes place with 20%, 50% and 80% training sets, each with 10 iterations, to obtain the average estimated risk. Table 5-19 to Table 5-21 lists the trained and tested values for each training set. The average estimated risk values for 10 random iterations for 20%, 50% and 80% training are 0.29, 0.326 and 0.097, which approximately correspond to 71%, 66% and 90% of the prediction. One exception that can be seen in the split-sample validation for the schema quality attributes is the relatively higher rate of predictability that has been achieved in the 80% training set. This rate is not only the highest among the different training sets for the schema quality attributes, but is also the highest among all the sample-split validation tests having 10 random iterations, for all scenarios. Nevertheless, the wider range of variation for the training set without any suggestive pattern may not necessarily result in higher confidence for the schema training with a high percentage of training in the dataset. Thus, an in-depth comparison of the different scenarios may be a more suitable approach to get a better insight instead of analysis of the case by case estimated risk. We discuss such a comparison of the validation of these three scenarios in the discussion in chapter 6.

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority			
				1	2	3	4
1	Training	0	0	Methodology	-	-	-
	Testing	0.429	0.187				
2	Training	0	0	Abstract	-	-	-
	Testing	0.235	0.103				
3	Training	0	0	Abstract	-	-	-
	Testing	0.4	0.110				
4	Training	0	0	Results	Abstract	-	-
	Testing	0.111	0.074				
5	Training	0	0	Introduction	-	-	-
	Testing	0.222	0.098				
6	Training	0	0	Abstract			
	Testing	0.529	0.121				
7	Training	0	0	Methodology	Abstract	-	-
	Testing	0.222	0.098				
8	Training	0	0	Introduction	-	-	-
	Testing	0.071	0.069				
9	Training	0	0	Abstract	Abstract	Abstract	-
	Testing	0.636	0.103				
10	Training	0	0	Abstract	-	-	-
	Testing	0.2	0.103				
Average	Training	0	0				
	Testing	0.29	0.1				

Table 5-19 Split-sample Validation of IGI's Content Quality Attributes with 20% Training

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority 1	Predictor Priority 2	Predictor Priority 3	Predictor Priority 4
1	Training	0	0	Methodology	Abstract	Conclusion & Discussion	-
	Testing	0	0				
2	Training	0	0	Methodology	Introduction	-	-
	Testing	0.182	0.116				
3	Training	0	0	Methodology	Introduction Abstract	Abstract	-
	Testing	0.538	0.138				
4	Training	0	0	Introduction	Abstract	-	-
	Testing	0.615	0.135				
5	Training	0	0	Methodology	Abstract Introduction	Conclusion & Discussion	-
	Testing	0.167	0.108				
6	Training	0	0	Methodology	Abstract	Abstract	-
	Testing	0.333	0.136				
7	Training	0	0	Methodology	Conclusion & Discussion	-	-
	Testing	0	0				
8	Training	0	0	Methodology	Introduction	-	-
	Testing	0.364	0.145				
9	Training	0	0	Methodology	Abstract	Abstract	-
	Testing	0.563	0.124				
10	Training	0	0	Methodology	Abstract	-	-
	Testing	0.5	0.144				
Average	Training	0	0				
	Testing	0.326	0.104				

Table 5-20 Split-sample Validation of Schema Quality Attributes with 50% Training

Iteration #	Training or Testing	Risk Estimate	Standard Error	Predictor Priority 1	2	3	4
1	Training	0	0	Methodology	Abstract Introduction	Conclusion & Discussion	-
	Testing	0.143	0.132				
2	Training	0	0	Methodology	Abstract Introduction	Conclusion & Discussion	-
	Testing	0.125	0.117				
3	Training	0	0	Methodology	Abstract Introduction	Abstract	-
	Testing	0.167	0.152				
4	Training	0	0	Results	Abstract Introduction	Abstract	-
	Testing	0.2	0.179				
5	Training	0	0	Results	Introduction	Methodology Abstract	Abstract
	Testing	0	0				
6	Training	0	0	Methodology	Conclusion & Discussion	-	-
	Testing	0	0				
7	Training	0	0	Methodology	Abstract Introduction	Abstract Conclusion & Discussion	-
	Testing	0	0				
8	Training	0	0	Results	Abstract Introduction	Abstract	-
	Testing	0.2	0.176				
9	Training	0	0	Methodology	Abstract Introduction	Abstract Conclusion & Discussion	-
	Testing	0	0				
10	Training	0	0	Methodology	Abstract Introduction	Abstract	-
	Testing	0.143	0.132				
Average	Training	0	0				
	Testing	0.097	0.088				

Table 5-21 Split-sample Validation of IGI's Content Quality Attributes with 50% Training

Figure 5-8 shows the prediction rate for different training rates in split-sample validation of schema quality attributes.

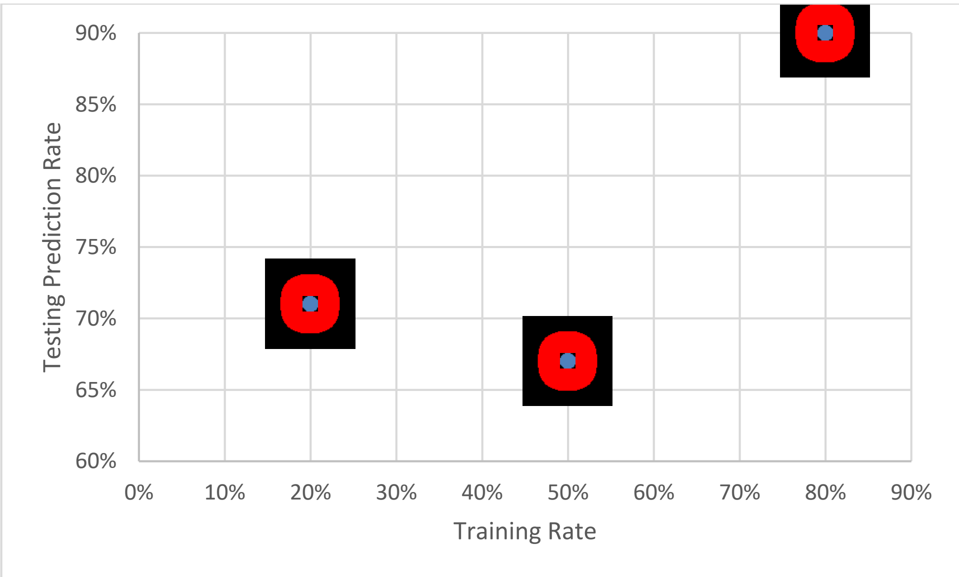


Figure 5-8 Split-Sample Validation Ratio for Schema Quality Attributes

5.5 Summary of Knowledge Quality Assessment Validation Analysis in Peer-Review

In this chapter, we discuss the validation methods by clustering the dataset obtained from the survey. The main reason behind such clustering is to see how peer-review decisions might be affected (in crossvalidation) by removing some reviewers from the sample size (decision scaling) or if the decisions can be replicated for another group of reviewers (split-sample validation). This chapter has also provided detailed analysis on the validation based on the scenarios that were developed in chapter 4. Such assessment of validation can be used to better assess the role and diversity of the knowledge quality attributes that are taken into consideration in order to arrive at a decision. It can give a better insight on whether the number of attributes or the number of people involved in a knowledge quality assessment process can affect the final decision and to what extent can the content and schema, as the main knowledge resources, affect decisions based on knowledge quality criteria.

In chapter 6, we discuss the results that were obtained in both chapter 4 and chapter 5 and feed these results into the research methodology developed in chapter 3. We then illustrate how the analysis in this study can contribute to the objectives of this research study.

Chapter 6 : DISCUSSION

In chapter 4 and chapter 5, we have extensive studied on decision trees and different types of validation related to them. While each method are illustrated briefly for each scenario, the main part of the discussion and interpretation of all those scenarios related to the three dimensions of Bias, Knowledge Quality and Decision Scaling is described in this chapter.

In the first part of this chapter, we investigate each of the three scenarios which are Emerald content quality attributes, IGI content quality attributes and schema content quality attributes separately. In the second part of this chapter, we look into the comparison of the content quality criteria obtained from IGI and Emerald and their characteristics. Finally, in the third part of this chapter, we discuss about the schema quality attributes and how they perform in comparison with IGI and Emerald on content quality variables.

6.1 Discussion on Content Quality Attributes Adopted from Emerald

In chapter 4, we initially discussed the descriptive data collected from the survey for all scenarios as the background of the participants, and how the decision has been affected by the different background information. We also discussed how decision trees are constructed, based on the grading of the quality attributes. We now interpret the collected data from the three dimensions of bias, knowledge quality and decision scaling.

6.1.1 Bias Dimension for Content Quality Attributes Adopted from Emerald

In this part, we are looking at the patterns of some of the backgrounds of the participants in relationship to each of the content quality variables that were adopted from Emerald.

In order to get a better insight about how the background information may affect the content quality attributes adopted from, we first look into the variables used to construct the decision tree for Emerald. In this scenario, the three variables of Result, Originality and Relationship to literature can best classify the dataset (See Figure 4-15). On the other hand, the participants have given their priorities among six choices of 1- Advancement of Knowledge, 2-Novelity of the proposed idea, 3- Validity of the proposed methodology 4-Realibility of the proposed hypothesis and research framework 5- Generalizability of the experiment and 6-Applicability of the research topic. In order to visualize the relationship between the impacts of the reviewer's priorities on his or her content quality grading, we summarize the means achieved by each priority through statistical indexing. Using statistical mean indexing for quality grading of high priority attributes (the ones that are in the decision tree) provides a better understanding rather than simply comparing each attribute separately. By using statistical mean indexing in the SPSS chart builder, we can get a summary for all the participating content attributes, which are 1- Result, 2-Originality and 3-Relationship to the literature review. Figure 6.1 shows a histogram of the statistical indexed mean of the three attributes used in the decision tree for this variable. As we have three attributes each ranging from 0 to 100, the statistical mean index for these three attributes will be between 0 to 300 (maximum 100 for each attribute). Figure 6-1 to Figure 6.6 shows the grading mean index in relation to each reviewer's content priority.

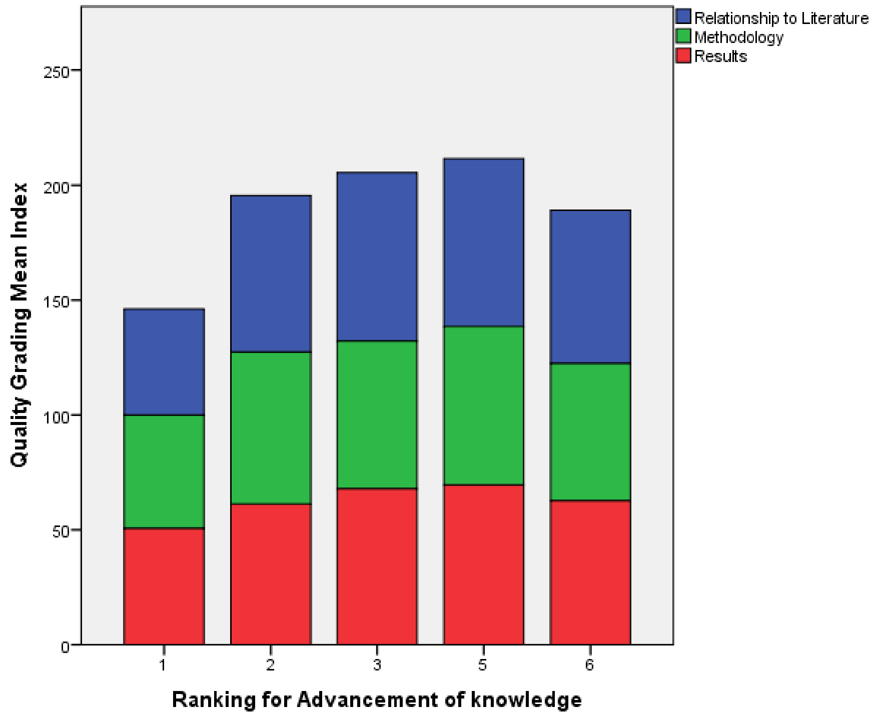


Figure 6-1 Content Bias Index for Advancement of Knowledge (Emerald Scenario)

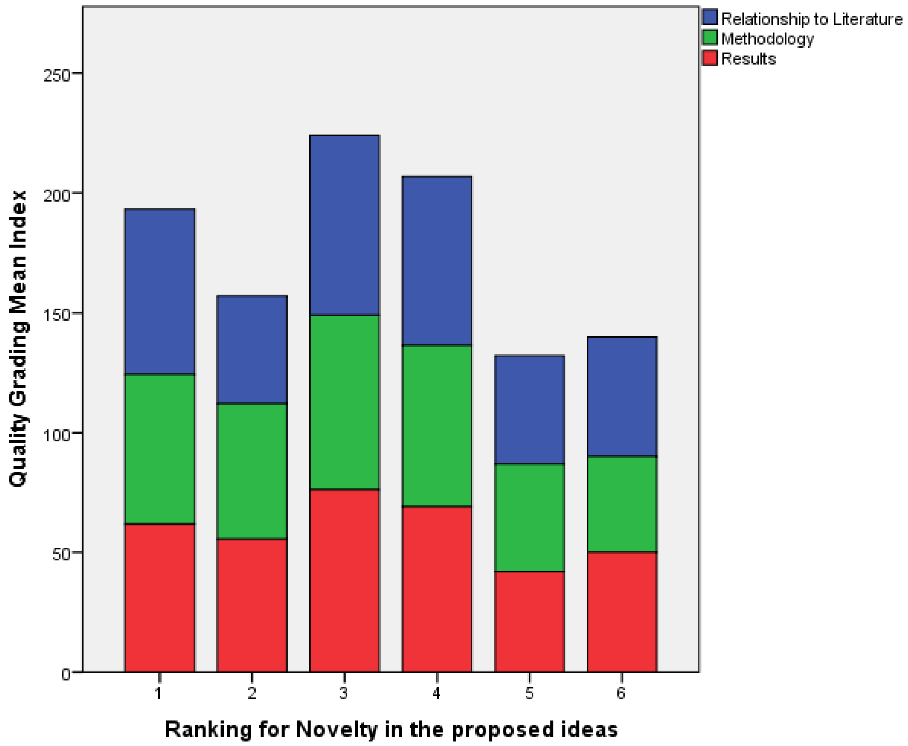


Figure 6-2 Content Bias Index for Novelty of Ideas (Emerald Scenario)

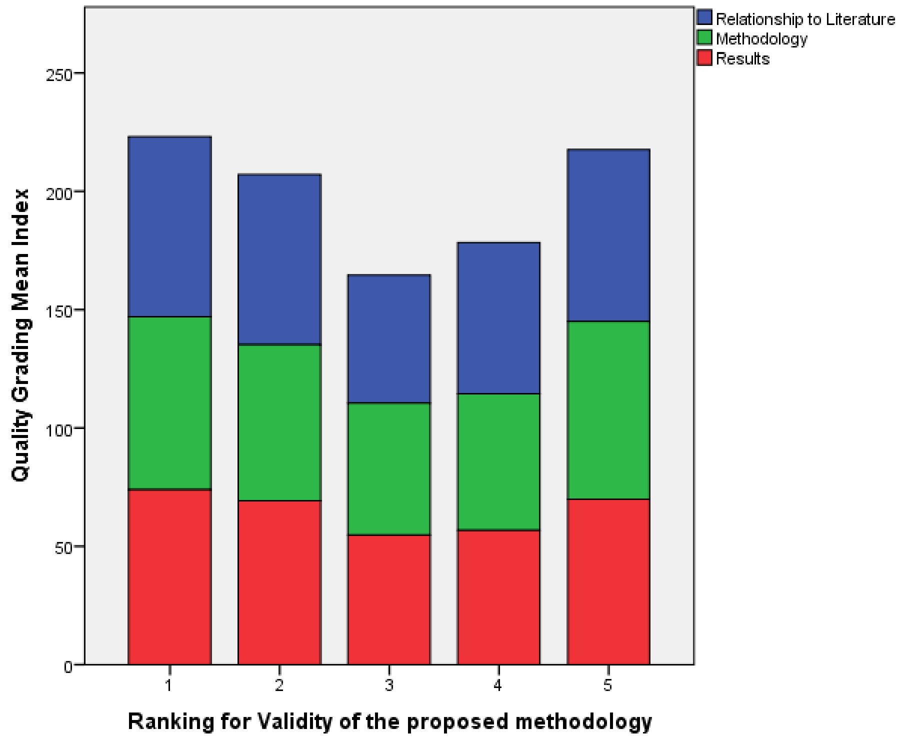


Figure 6-3 Content Bias Index for Validity of Methodology (Emerald Scenario)

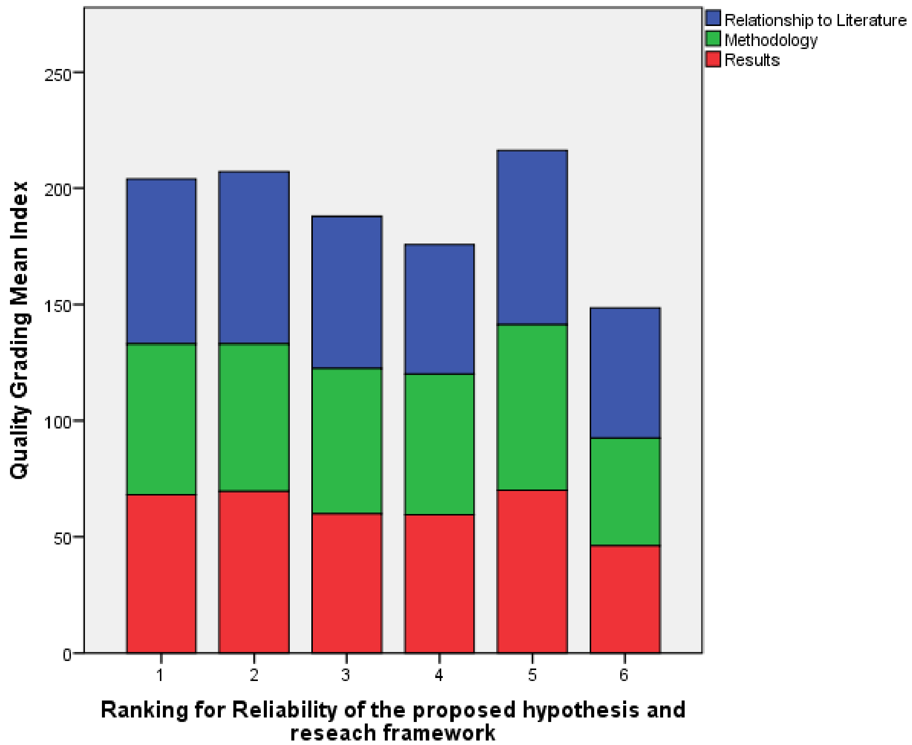


Figure 6-4 Content Bias Index for Reliability of Methodology (Emerald Scenario)

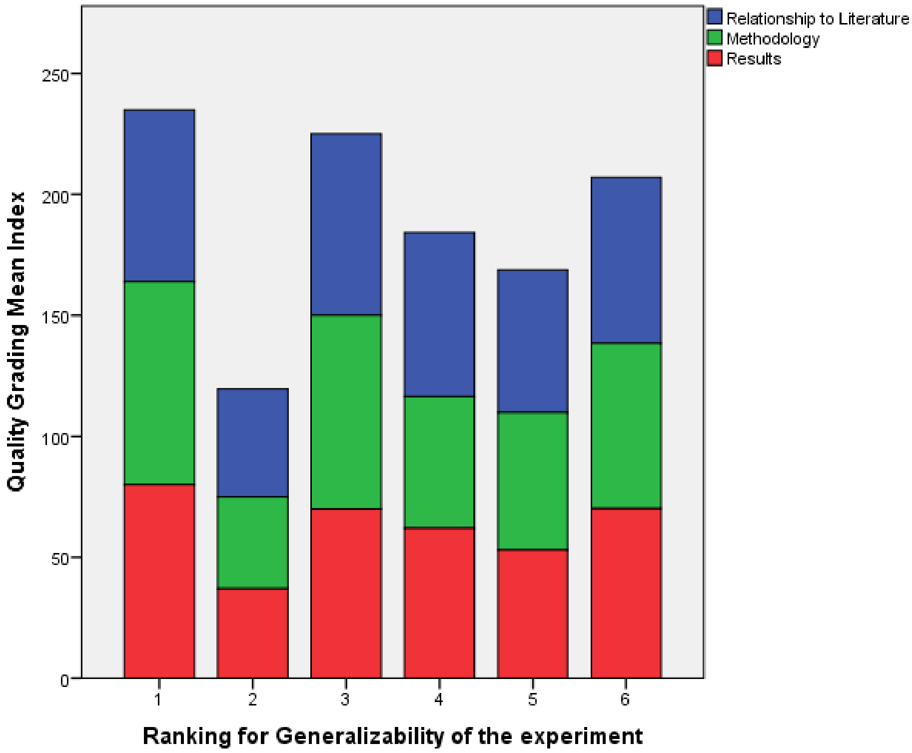


Figure 6-5 Content Bias Index for Generalizability (Emerald Scenario)

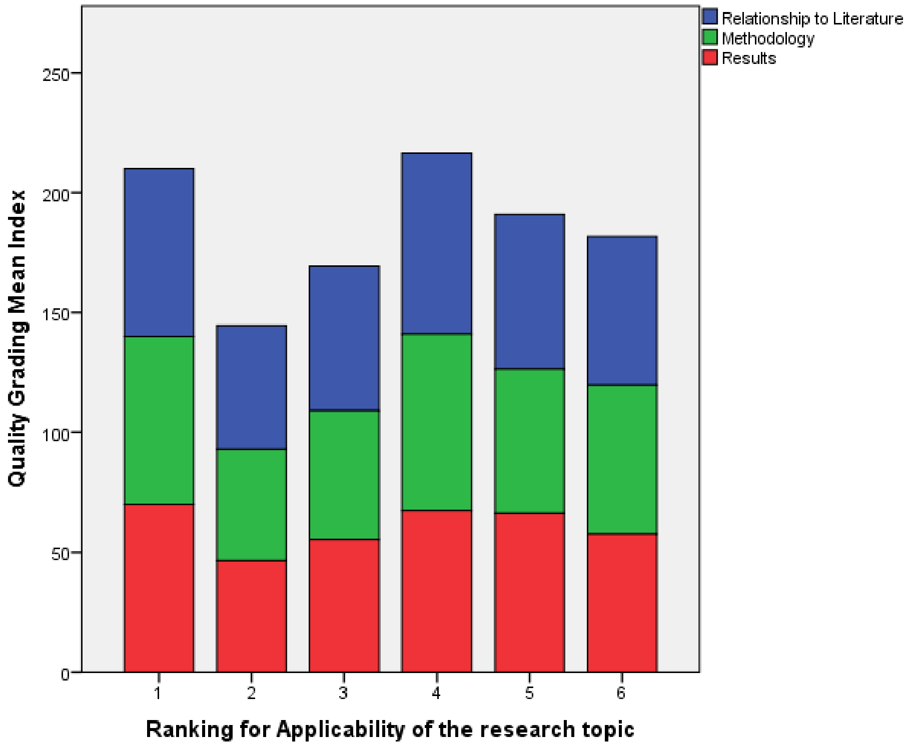


Figure 6-6 Content Bias Index for Applicability (Emerald Scenario)

As it can be seen from Figure 6-1 to Figure 6-4, different priorities can achieve different means from the quality grading index, in association with content quality attributes, as good predictors in their respective decision tree. The higher variation the bar levels have across priorities, the stronger is the role they can play in the decision path. As an example, based on the quality content variables adopted from Emerald, the generalizability of the experiment, to be the first or second preference, has a significant impact on the quality grading mean index, thus, can take the decision path in a very different direction.

6.1.2 Knowledge Quality Dimension for Content Quality Attributes Adopted from Emerald

Out of the total number of six content quality attributes that were adopted from Emerald, only three were used in the decision tree for 100% of the cases. The validation test for this scenario that was illustrated in chapter 5, showed that the three variables of Results, Originality and Relationship to Literature could also classify decisions for other decision population subsets, to various degrees. As shown earlier in chapter 5, these three variables could classify smaller chunks of the dataset through folding. Table 6-1 shows a summary of the prediction (classification) percentage obtained for different folds from these three variables in 10 iterations. As it can be seen, from these three attributes, 70% to 73% of the cases in different subset of the database can be predicted. This can be interpreted, with only these three variables, that the majority of the other cases can also be predicted. In other words, the other attributes strongly follow the same pattern of these three variables (high level of correlation) or they cannot play a very important rule in the knowledge quality assessment decision making.

Number of Folds	Percentage of Prediction	Average Estimated Error
2	73%	0.084
3	70%	0.086
4	71%	0.086
5	73%	0.087

Table 6-1 Summary of Content Quality Prediction Using Crossvalidation (Emerald Scenario)

The other type of validation is the split-sampling validation that can train a subset of the dataset and apply it to the testing set. This validation can basically could reconstruct the decision trees on different sets which are a random subset of the total population (27 participants). With the split-sample validation test, we can count the number of occurrences of each of the content quality variables for different subsets. This can help us to understand the chance of predictability and importance of each of the content quality attributes that were adopted from Emerald. Table 6-2 shows a summary of the occurrences of these content quality variables for 10 iterations in each randomly chosen subset from the whole population.

Content Quality Attribute	Training Percentages & Number of Occurrences (out of 10)			
	20%	50%	80%	Total (Out of 30)
Originality	7	9	10	26
Relationship To Literature	0	2	5	7
Methodology	1	2	4	7
Results	3	4	7	15
Implications for research, practice and/or society	1	0	0	1
Quality of Communication and Language	0	0	0	0
Average Prediction	65%	71%	68%	

Table 6-2 Summary of Content Quality Prediction Using Split-Sample (Emerald Scenarios)

Table 6-2 shows that "Originality" is present in 26 cases out of 30 for different training percentages. With the highest number of occurrences in this quality attribute set, adopted from emerald, it shows its key role in any derivation of a randomly chosen subset from the studied population. The attribute "Results" comes second followed by "Relationship to Literature Review" and "Methodology". Hence, the quality attributes of Originality and Results are the best predictors in any size of subset

6.1.3 Decision Scale Dimension for Content Quality Attributes Adopted from Emerald

The last dimension considered for the Emerald scenario is the Decision Scale Dimension. From Chapter 5, with both crossvalidation and split-sampling, we clustered, trained and tested the dataset to see how it responds to various sizes for knowledge quality assessment decision making. As Table 6-3 shows, both crossvalidation and split-sample validation do not suggest that the peer-review decision pattern may be very sensitive to scaling. The random sampling based on the content quality attributes adopted from Emerald do not suggest that the growth in sample size affects the prediction in a stable manner. It should also be noted that in case of the split-sample at 20%, due to the small size of the sample, the prediction level might not be stable when there are only a few cases for training.

Test Type	Fold or Train%	Estimated size of fold or training set	Worst Prediction	Best Prediction	Average Estimated Prediction	Average standard Error
Crossvalidation	2 Folds	14 cases	61%	75%	73%	8%
Crossvalidation	3 Folds	18 cases	63%	85%	70%	9%
Crossvalidation	4 Folds	20 cases	63%	78%	71%	9%
Crossvalidation	5 Folds	22 cases	67%	78%	72%	9%
Split-sample	20% Training*	5 cases	21%	88%	65%	16%
Split-sample	50% Training	14 cases	54%	92%	71%	12%
Split-sample	80% Training	22 cases	50%	100%	67%	16%

Table 6-3 Summary of Decision Scale Prediction Using Validation (Emerald Scenarios)

6.2 Discussion on Content Quality Attributes Adopted from IGI

Similar to the attributes adopted from Emerald, a series of attributes were also adopted from IGI for evaluation. The major difference in these two scenarios on the content attributes (other than variation in wording and definition) concerns the number of attributes. Adopted attributes from IGI had totally 9 attributes while for Emerald, the total number was 6. In this part, we discuss the analysis of these 9 variable and compare it against Emerald scenarios update in the previous section.

6.2.1 Bias Dimension for Content Quality Attributes Adopted from IGI

The IGI scenario has totally nine attributes: 1- Popularity of the Subject 2- Appropriateness for the journal 3- Adequacy of literature review 4- Quality of research design 5- Adequacy of data analysis 6- Contribution to the literature 7-Legitimacy of the conclusion 8-Practical/managerial significance and 9-Clarity of presentation. Out of these nine content quality variables, only three are involved in constructing the decision tree (Fig 4-19). These three variables are Adequacy of Data Analysis, Popularity of the Subject and Quality of Research Design. This means, with having more content quality attributes in the knowledge quality evaluation, we still have the same number of predictors that can classify 100% of the cases. Similar to the Emerald scenario, Figure 6-7 to Figure 6-12 show the pattern of the reviewer's preferences for these three content quality variables by generating a mean index. Similarly, with three variables, the quality grading index will also be between 0 to 300.

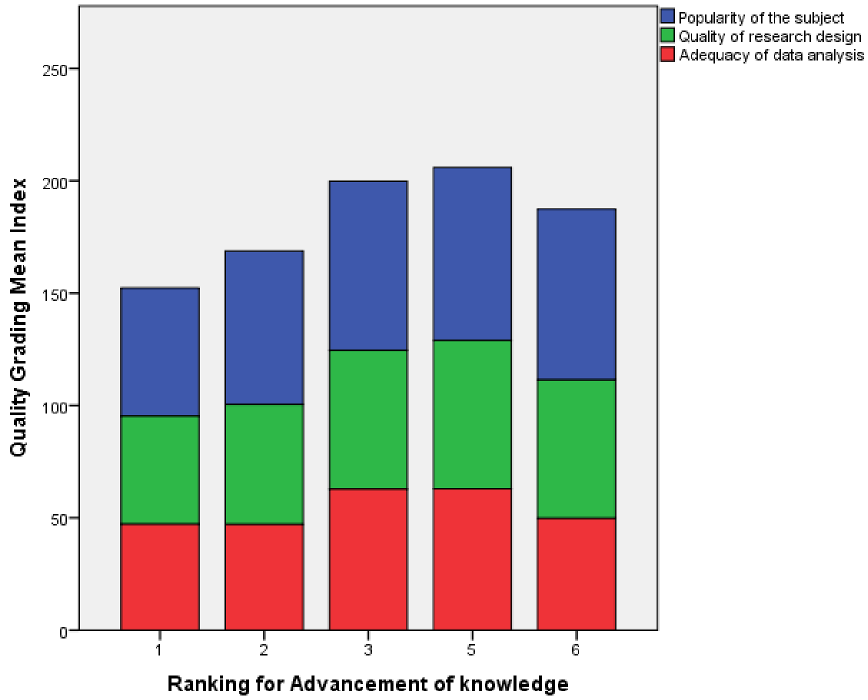


Figure 6-7 Content Bias Index for Advancement of Knowledge (IGI Scenario)

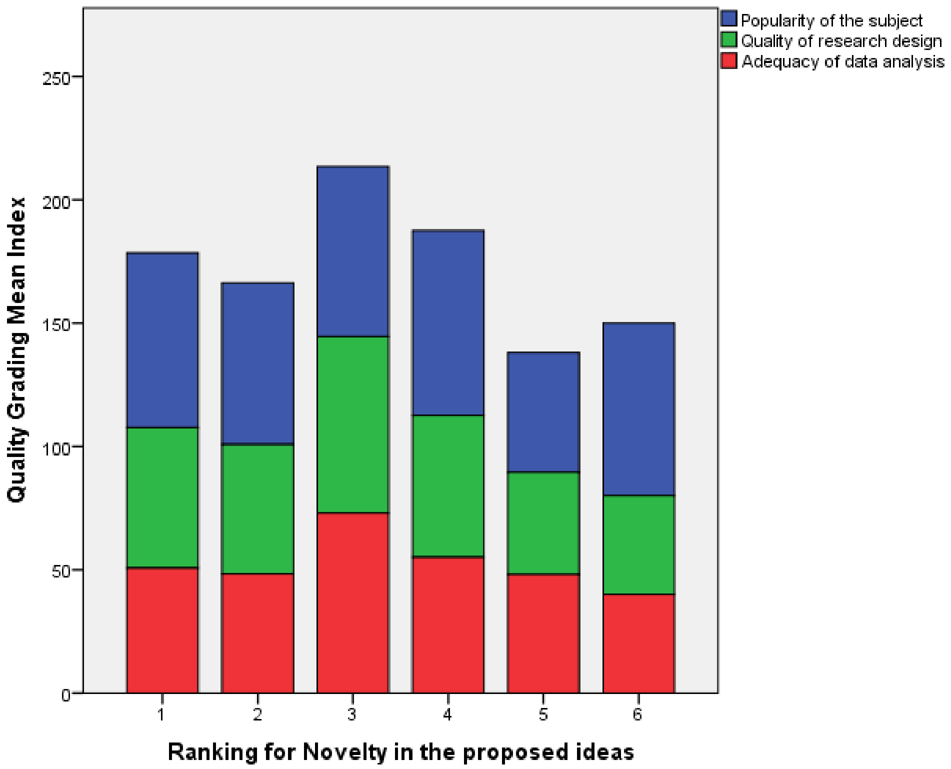


Figure 6-8 Content Bias Index for Novelty of Ideas (IGI Scenario)

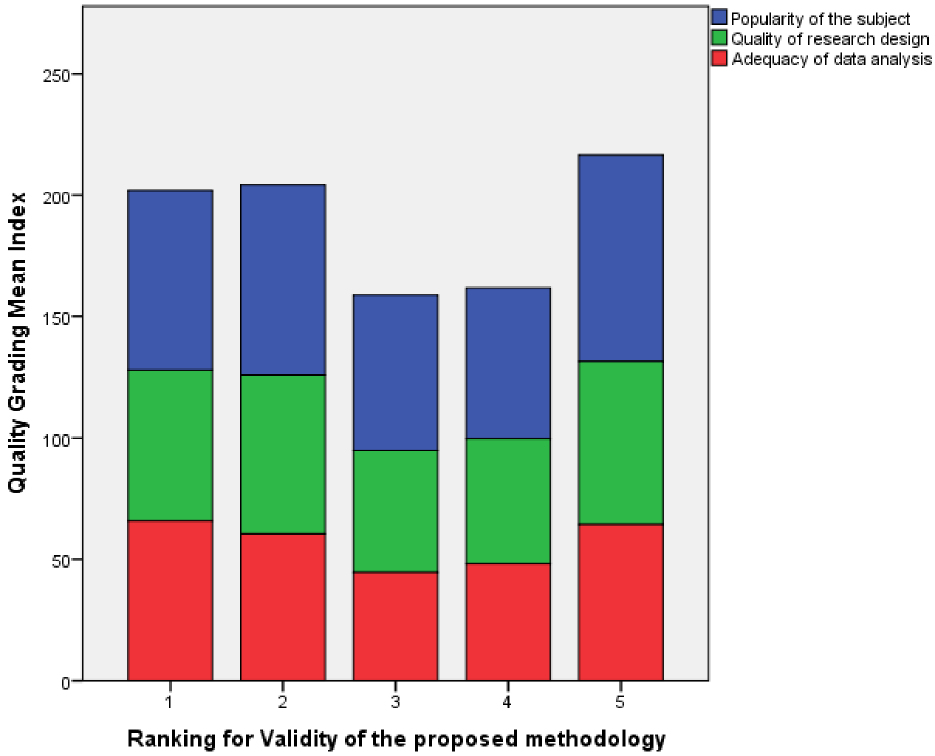


Figure 6-9 Content Bias Index for Validity of Methodology (IGI Scenario)

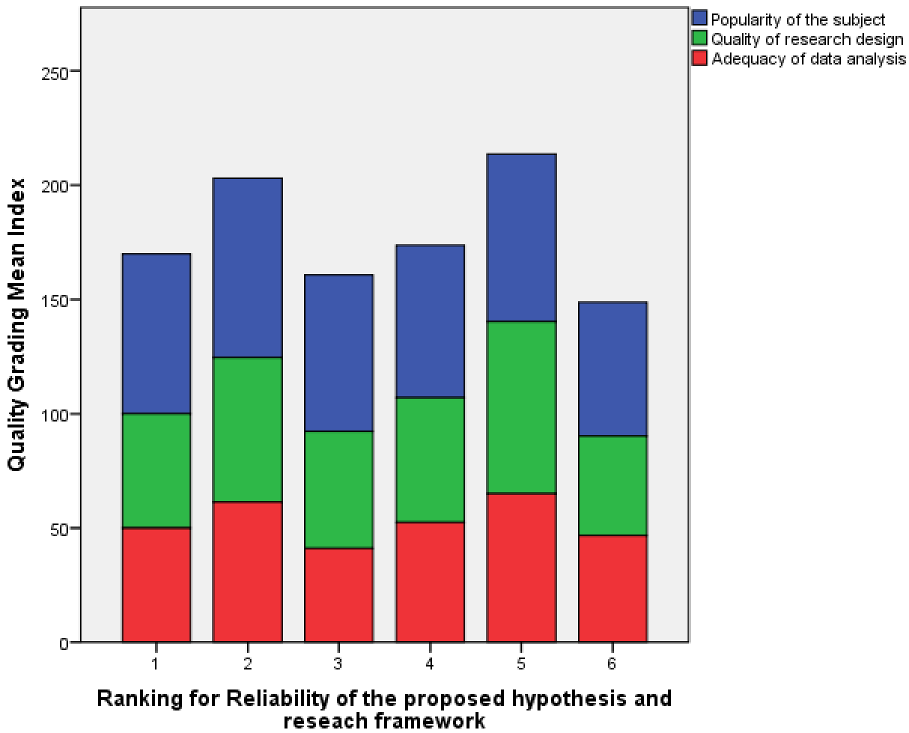


Figure 6-10 Content Bias Index for Reliability of Methodology (IGI Scenario)

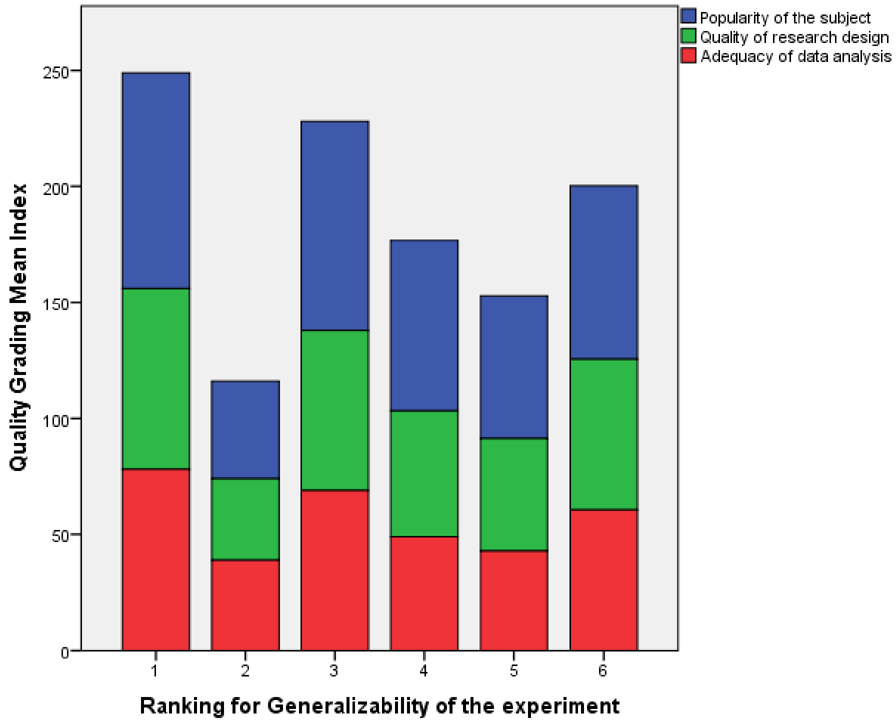


Figure 6-11 Content Bias Index for Generalizability (IGI Scenario)

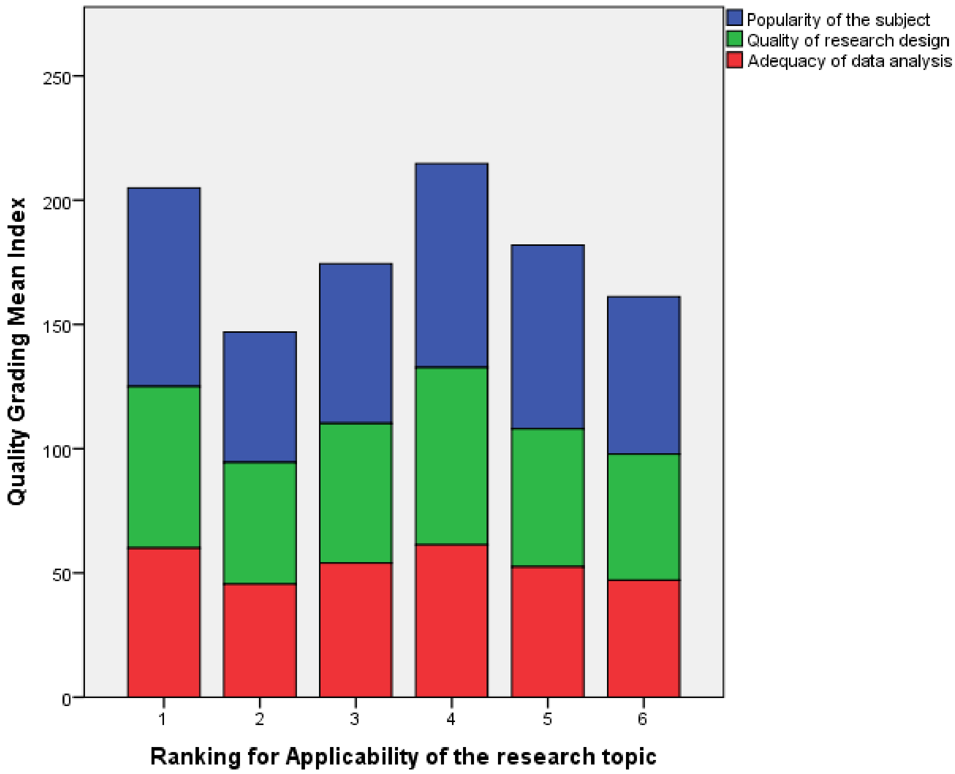


Figure 6-12 Content Bias Index for Applicability (IGI Scenario)

As Figures 6-7 to 6-12 show, similar to the Emerald case, there are variations in the index gain for different priorities within different preferences. Nevertheless, the level of variation maybe significant among some priorities while it may not show a big variation in some other cases.

In order to get a better insight about these preferences, comparing the two scenarios of IGI and Emerald may provide a bigger picture of a reviewer's preference impact on the decision tree paths. Figure 6-13 summarizes all the histograms obtained from the both scenarios and the pattern of changes of the quality grading mean index for all the six preferences, based on the content quality attributes of the decisions tree for both Emerald and IGI. As can be seen, the summary of changes in the preference priorities, when compared, shows many pattern similarities for both sets of variables. While there are slight variations in each scenarios index gain over a preference priority, nevertheless, the general pattern for quality grading over different sets of content quality variables follows relatively similar patterns. As the content quality attributes are the main predictors of the decision tree, a big gap between the gains of the bar for these predictors can suggest a different decision path (irrespective of the decision outcome). One good example is the preference for generalizability of experiments which shows a huge difference for both scenarios between being the first or the second priority. With the difference of mean index being more than 100, it can suggest, using the three predictors in each scenario, that the reviewer's decision path is different when considering generalization of experiments as the main point of focus and preference. Applicability of the research topic might be another preference which shows a similar pattern between being the first or the second preference.

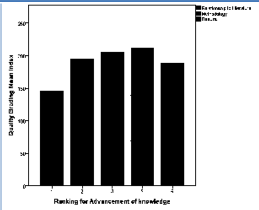
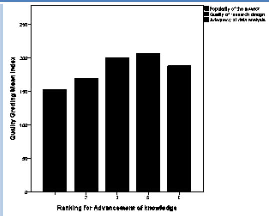
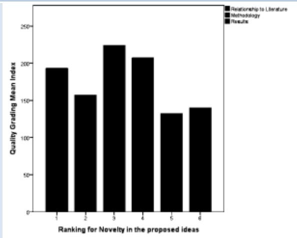
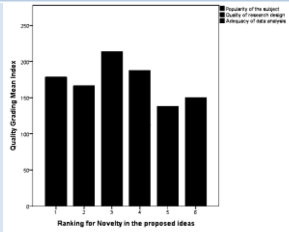
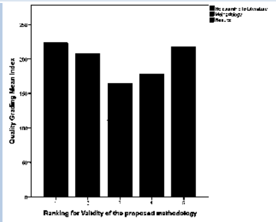
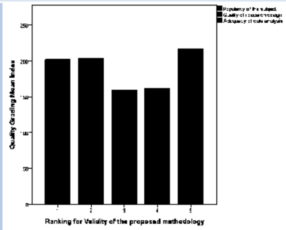
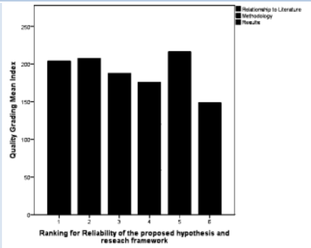
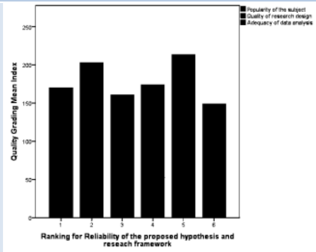
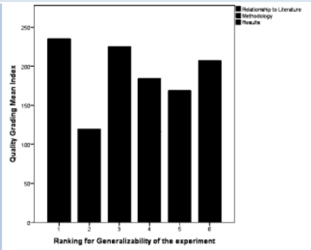
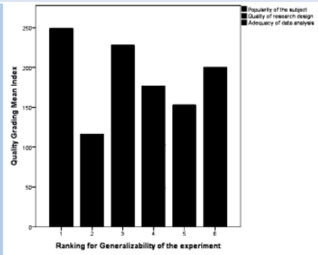
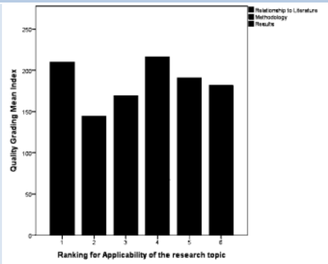
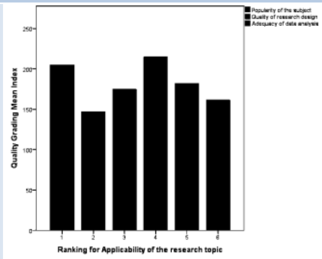
Reviewer's Content Preference	Emerald	IGI
Advancement of Knowledge		
Novelty of the proposed idea		
Validity of the proposed methodology		
Reliability of the proposed hypothesis and research framework		
Generalizability of the experiment and		
Applicability of the research topic		

Figure 6-13 Comparison of Preference-Quality Grading Mean Variation

The other point that can be drawn from Figure 6-13 is the relatively high similarity of the patterns for the quality grading mean index between the two scenarios. While the preferences stay the same for each reviewer, the two series of content quality attributes to grade are different in wording and definition. However, such similarity in the pattern for both scenarios from the adopted content quality attributes shows that the quality grading index patterns for the same reviewer with defined preferences won't alter so much in regard to the way he/she acts in assessing the knowledge quality. In other words, the reviewers are not too sensitive over variations in content quality attributes which may suggest that the reviewers cannot fully differentiate all the content quality attributes and consider the knowledge quality assessment for the content holistically based on their preferences and understanding. This can also justify the very high correlation levels between different sets of content quality attributes and the omission of a big portion of the content quality attributes when constructing the decision trees. This is because, out of too many quality attributes related to the content, a reviewer may only be able to recognize and differentiation a few of them. As has been seen, the increased number of content quality attributes in the IGI case compared to Emerald did not add any quality attributes to the structure of the decision tree. We explore other aspects of the number of variables in the section on decision scaling comparison.

6.2.2 Knowledge Quality Dimension for Content Quality Attributes Adopted from IGI

In chapter 5, we also did similar crossvalidation and split-sample validation tests for the content quality attributes that shape the decision tree in the IGI scenario. Same to the Emerald case earlier in this chapter, we further investigate these content quality attributes for the IGI scenario.

In the first validation, which was crossvalidation, we applied the same folding option to both the IGI and Emerald scenarios. Table 6-4 shows a summary of the crossvalidation and predictability of the content quality variables for the IGI decision tree. These variables, as mentioned earlier, are 1- Adequacy of data analysis 2- Popularity of the subject and 3- Quality of research design.

Number of Folds	Percentage of Prediction	Average Estimated Error
2	73%	0.083
3	78%	0.079
4	76%	0.081
5	75%	0.082

Table 6-4 Summary of Content Quality Prediction Using Crossvalidation (IGI Scenario)

The prediction variables from IGI are slightly better than the ones in Emerald for decision prediction in 10 iterations for each fold. While the prediction average of all the folds together for Emerald is approximately 73%, the average for IGI is around 75.5%. One interpretation for this slightly better performance might be the increased number of variables from 6 (Emerald) to 9 (IGI) in the quality grading assessment. Although for both scenarios, they are reduced to same size of three variable in their decision trees as the best predictors; nevertheless, more variables for the small sample size in this study may end up with choosing better predictors. As mentioned earlier, with a reviewer's lack of sensitivity over the content quality criteria for assessment, even a slightly better performing predictor may help to increase the average outcome in the classification of the decision path. Such slight variation may also be the result of overfitting which is discussed in chapter 7 as the limitation on small sample size.

The other validation test is split-sampling using the training and testing sets. For the IGI scenarios, Table 6-5 shows a summary of tests described in chapter 5.

Content Quality Attribute	Training Percentages & Number of Occurrences (out of 10)			
	20%	50%	80%	Total (Out of 30)
Popularity of the Subject	7	10	10	27
Appropriateness for the journal	2	3	3	8
Adequacy of literature review	3	0	0	3
Quality of research design	1	5	8	14
Adequacy of data analysis	0	6	9	15
Contribution to the literature	0	1	1	2
Legitimacy of the conclusion	0	0	0	0
Practical/managerial significance	0	0	0	0
Clarity of presentation.	0	0	0	0
Average Prediction	66%	72%	69%	

Table 6-5 Summary of Content Quality Prediction Using Split-Sample (IGI Scenarios)

As it can be seen from Table 6-5, the bigger the training set, the more frequent the number of occurrences for the predicting variables. Nevertheless, the average prediction in each stage for the IGI scenarios is not much different from Emerald, and also there are not very big variations among the different training sets.

Based on both scenarios of IGI and Emerald, the performance of the quality attributes that shape their decision trees are relatively similar, with only minor variations that can be slightly better for IGI due to larger pool of content quality variables to select.

6.2.3 Decision Scale Dimension for Content Quality Attributes Adopted from IGI

In the last part of the discussion on and comparison of the content quality attributes, the decision scale is investigated for the IGI scenario and is compared to the Emerald scenario described earlier in this chapter.

By combining both the crossvalidation and split-sample validation results obtained from the IGI scenario, we look into how scaling the decision size, based on the variables in the IGI decision tree, affects the predictability of the reviewer's decision. Table 6-6 summarizes the predictions for each fold and training percentage. While, on average, the prediction level here is slightly better in comparison with the Emerald scenario, nevertheless, the difference in the best and worst prediction is higher for IGI. Hence, it results in higher standard error. The pattern of diversity for the average prediction is similar to Emerald, but the decision scaling risk is higher due to the higher standard error. As crossvalidation generally performs better for the IGI scenario while there's a higher risk for the split-sample validation, it can be interpreted that the IGI adopted variables work better for the designated dataset obtained from the survey. However, they may not perform very well for scaling, due to overfitting which is further discussed in chapter 7.

Test Type	Fold or Train%	Estimated size of fold or training set	Worst Prediction	Best Prediction	Average Estimated Prediction	Average standard Error
Crossvalidation	2 Folds	14 cases	59%	85%	73%	8%
Crossvalidation	3 Folds	18 cases	70%	82%	78%	8%
Crossvalidation	4 Folds	20 cases	67%	89%	76%	8%
Crossvalidation	5 Folds	22 cases	67%	82%	75%	8%
Split-sample	20% Training*	5 cases	43%	93%	66%	11%
Split-sample	50% Training	14 cases	43%	100%	72%	11%
Split-sample	80% Training	22 cases	34%	100%	69%	19%

Table 6-6 Summary of Decision Scale Prediction Using Validation (IGI Scenarios)

6.3 Discussion on Schema Quality Attributes

The last part of this chapter analyses the results obtained from preferences and quality grading used for the schema (format) by the reviewers. In the schema, the focus is on the format and structure of the manuscript that is usually observed as the compliance or monitoring elements of the manuscript review. Similar to the content quality factors, we investigate the three dimensions of bias, knowledge quality and decision scale, but with the preferences and attributes that are based on schema rather than related to the content.

6.3.1 Bias Dimension for Schema Quality Attributes

In this part, we look into the preferences that were listed in the survey and explained in chapter 4 (Table 4-11). These preferences are 1- Abstract 2- Introduction to

Literature review 3- Research Framework and Methodology 4- Discussion and Conclusion and 5- References. There were also a total of 8 schema quality variables that were obtained from the participants to grade the selected manuscript based on the structure and format (table 4-16). These eight schema variable are 1- Abstract 2- Introduction 3-Methodology 4- Results 5-Conclusion and Discussion 6- Adequate References 7- Accurate References 8- Up-to-date references. Out of these eight variables, the four variables “Abstract”, “Methodology”, “Introduction” and finally “Conclusion and discussion” predicted 100% of the dataset in decision tree as discussed in chapter 4 (Figure 4-23). Here, similar to previous cases in this chapter for bias dimension analysis, we create a mean index from the participating variables in decision tree to see how the schema preference priorities may vary for the total schema quality grading mean index. With four variables here, each ranging from 0 to 100, we have an index of 0 to 400 for the schema quality mean index. Figure 6-7 to 6-11 show how these variables can gain a mean index for each priority. As can be seen, the variation of the histogram for priorities in Research Framework and Methodology is higher than other preferences. The preferences for abstract and references do not lose or gain much over their priorities. The preference of introduction and literature review, once becoming the second priority or lower, will gain in the mean index overall. In summary, by looking into the mean index, these bias pattern for the research framework and methodology shows a more significant change of the quality mean index for the related variables. On the other hand, preference for the abstract is relatively constant for different priorities.

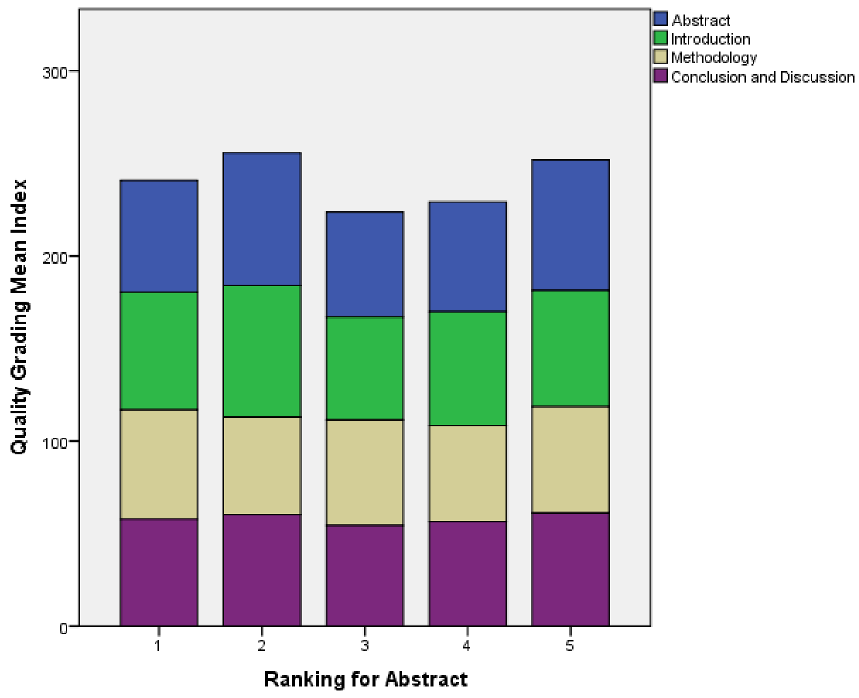


Table 6-7 Schema Bias Index for Abstract Preference

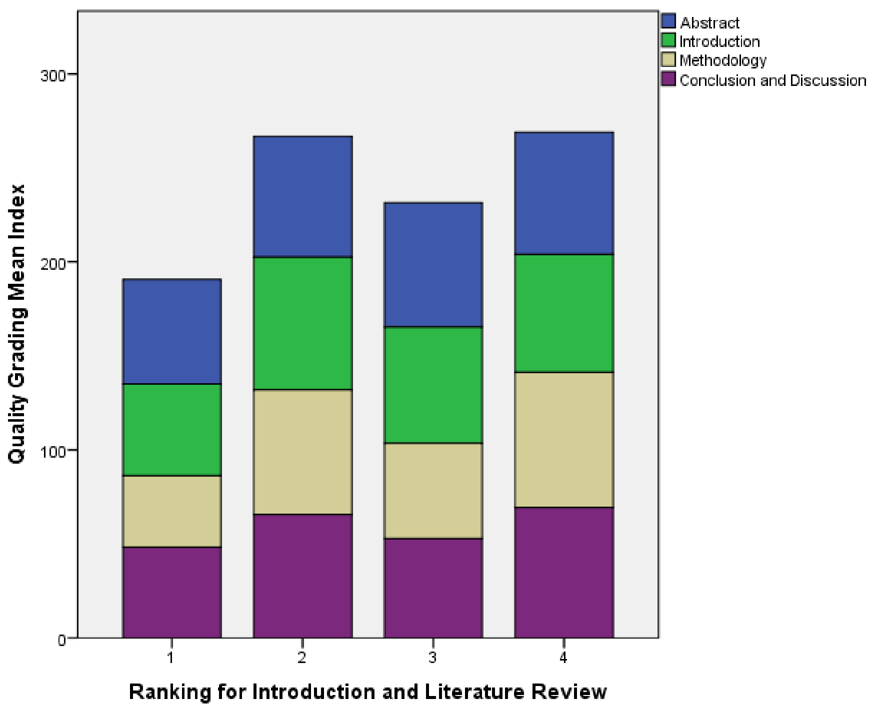


Table 6-8 Schema Bias Index for Introduction & Literature Review

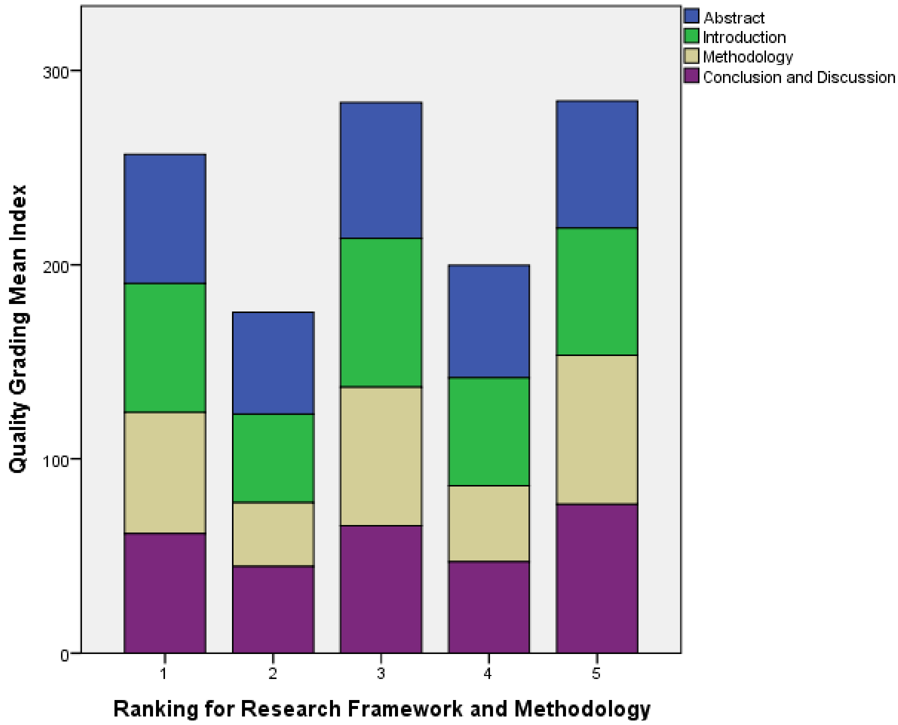


Table 6-9 Schema Bias Index for Research Framework & Methodology

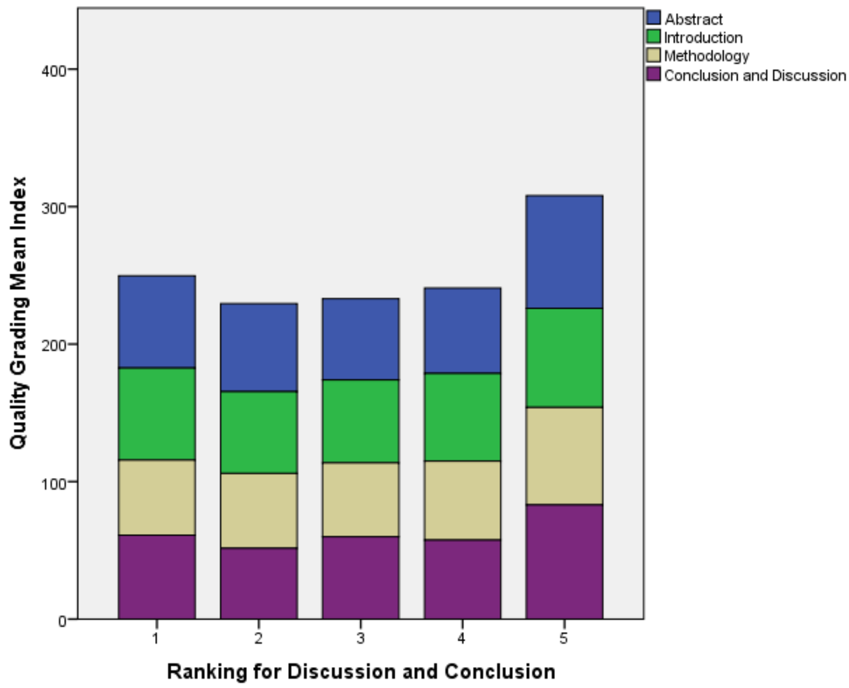


Table 6-10 Schema Bias Index for Discussion & Conclusion

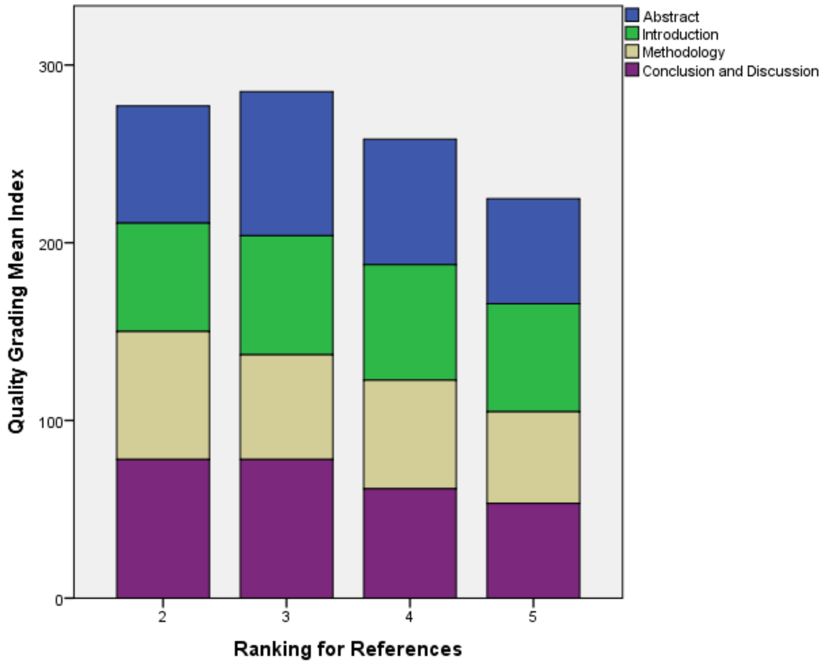


Table 6-11 Schema Bias Index for References

6.3.2 Knowledge Quality Dimension for Schema Quality Attributes

Here, we discuss the knowledge quality dimension related to structure and format. Similar to the content quality attributes, we also did both crossvalidation and split-sample validation for the predictably level of quality attributes related to the format of a manuscript. Table 6-12 shows a summary of the crossvalidation for schema related quality attributes. The percentage of prediction for different folds is between 74% to 78%. Based on the obtained values from the 10 iterations, the prediction performance in crossvalidation is relatively similar to the average prediction level we obtained from the content quality scenarios, being especially close to the IGI scenario on values and standard error. Thus, the schema quality variables are almost performing in a similar fashion to the quality content variable analysed in this dataset. It also classifies all the cases of the given dataset using the CRT growth method similar to the content quality scenarios.

Number of Folds	Percentage of Prediction	Average Estimated Error
2	74%	0.081
3	75%	0.082
4	79%	0.060
5	78%	0.070

Table 6-12 Summary of Schema Quality Prediction Using Crossvalidation

The other validation test reported in chapter 5 for the schema quality variables is the split-sample method. Table 6-13 shows a summary of the split-sample validation. This summary shows that the training performance for the schema quality attributes, after the size of training exceeds 50%, performs better than the content quality attributes. Nevertheless, in order to understand the real performance of the schema quality attributes, it may need to be trained and tested on larger datasets with a high number of iterations. At the level of the dataset used in this study, the training on the schema with larger subsets (80%) showed that on average, the prediction exceeds the content quality attributes by doing more iterations to achieve more confidence levels.

The other points to be seen in Table 6-13 is the quicker adjustment of the high performance predicting quality attributes for the construction of the decision tree with the expansion of the training set. While, for the content, small training sets may indicate the predicting of variables which will be then dropped from the larger training sets, the schema quality attributes are slightly more consistent over the training set expansions.

Schema Quality Attribute	Training Percentages & Number of Occurrences (out of 10)			
	20%	50%	80%	Total (Out of 30)
Abstract	7	6	9	22
Introduction	2	4	6	12
Methodology	2	9	8	19
Results	1	0	3	4
Conclusion and Discussion	0	2	4	6
Adequate References	0	0	0	0
Accurate References	0	0	0	0
Up-to-date references	0	0	0	0
Average Prediction	71%	68%	90%	

Table 6-13 Summary of Schema Quality Prediction Using Split-Sample

6.3.3 Decision Scale Dimension for Schema Quality Attributes

In the last part of the analysis in the discussion chapter, we will investigate the decision scaling, based on the schema quality attributes. In order to do this, similar to the content quality attributes, we combine the outputs from both crossvalidation and split-sample validation. Table 6-14 shows a summary of the two validation methods combined. Comparing the results with that obtained from the content quality attributes, the schema quality attributes show a better performance in predicting the decisions though scaling. It has also a faster convergence to the right classification, with expansion of the training set or addition of the cases to the dataset.

Test Type	Fold or Train%	Estimated size of fold or training set	Worst Prediction	Best Prediction	Average Estimated Prediction	Average standard Error
Crossvalidation	2 Folds	14 cases	52%	83%	74%	8%
Crossvalidation	3 Folds	18 cases	67%	83%	75%	8%
Crossvalidation	4 Folds	20 cases	70%	85%	79%	6%
Crossvalidation	5 Folds	22 cases	70%	83%	78%	7%
Split-sample	20% Training*	5 cases	47%	92%	71%	10%
Split-sample	50% Training	14 cases	49%	100%	67%	10%
Split-sample	80% Training	22 cases	80%	100%	90%	8%

Table 6-14 Summary of Decision Scale Prediction for Schema Using Validation

6.4 Summary of Discussion Chapter

Based on the analysis and discussion of all the scenarios in this chapter, the schema quality attributes show better performance for all the three dimensions. For the reviewer bias dimension, while both the content and schema show variation, such variation is more limited in schema assessment when compared to the content quality assessment. In the knowledge quality dimension, based on the validation results, it can be seen that a reviewer has a better sense of differentiating and grading the schema, so that the predicting variables converge faster and in a slightly more consistent way that in the care of the content. In the decision scaling dimension (inter-subjectivity), the schema quality attributes show better results for the smaller subsets (folds) or training sets (with lower percentage).

In the next chapter, which is the final chapter for this study, we summarize all the analyses and discussion based on the proposed conceptual framework. We also talk about the limitation of this study and how these limitations can be overcome in future research works.

Chapter 7 : CONCLUSIONS AND FUTURE RESEARCH

This research has investigated the assessment of quality of knowledge in the context of the peer-review process. Through the exploratory investigation in this study, some of the important characteristics of knowledge quality assessment have been highlighted. The role of content and schema as knowledge resources were also explored in detail. In this chapter, we summarize all of the findings illustrated and discussed during this study. We also discuss the significance and contribution of this study by highlighting the importance of a proper understanding about knowledge quality and its assessment. We also consider the limitations of this study in terms of dataset input and analysis, and propose solutions for such limitations in future research.

7.1 Findings

In this study, we explored the knowledge quality assessment by proposing three different scenarios within the context of the peer-review process. We also chose a domain of scientific knowledge, Knowledge Management, for this study in order to find more accurate quality measurement sources. With the schema (format) and the content, as the two main knowledge resources that should be taken into consideration in quality of knowledge evaluation, this study looked into the quality assessment criteria that are most commonly used for knowledge quality assessments in the peer-review process. In order to do so, we adopted the quality assessment criteria from the two well-known publishers in the Knowledge Management discipline to evaluate the decision patterns arising from the quality assessment attributes. In doing so, we proposed three different scenarios. The first two scenarios focused on content quality assessment which were adopted from the quality assessment criteria used in Emerald and IGI-global. The third scenario was built on the schema (format) requirements that are requested and adopted by the major publishers.

The data collected and analysed in this study were analysed with a descriptive approach. The descriptive statistical part reflected the summary of data collected throughout the survey and analytical decision tree methods were used to mine the decision patterns originating from the collected dataset. Thus, the findings of this study are divided into two main categories. The first one which is the output of the descriptive statistical findings and the second is on the findings from the analytical methods obtained from decision tree modelling and mapping.

7.1.1 Descriptive Statistical Findings

In the first part of chapter 4, we described the descriptive statistical data collected through the survey in this study. The focus of the descriptive statistical analysis was on the background information of the reviewers and the distribution of the peer-review decisions from those backgrounds.

One of the findings that could be observed from this research study were the experience, expertise and academic position of the reviewers. The diversity of decisions/opinions were seen at both ends of expertise and experience, where the diversity of the decisions for junior and senior participants were more than those with a few years of experience. This suggests that the inter-rater reliability which causes disagreement among scholars on the quality of a research work tends to be lower at the boundaries of the expertise level while the core population at the center of a scientific field think, relatively in a more similar manner. Moreover, it is seen that reviewers who have no expertise in the domain that they are being asked to do a review, tend to have more diversified decisions. Thus, this study showed that the inter-rater reliability of decisions on a manuscript diverges when the reviewers move to the boundaries of a scientific domain, both in terms of expertise and academic ranking.

7.1.2 Analytical Findings from the Quantitative Study

The analytical findings of this study discussed the quality knowledge decision patterns from three different dimensions: 1- Human Bias 2-Knowledge Quality Attributes and 3-Decision Scaling, which originated from either the schema or content.

From a human bias perspective, reviewers showed more diversified bias on the content than that of the schema. What has been seen as a result of the decision tree analysis is the very strong correlation between the grading outputs of the content variables. This can be interpreted as the inability of the reviewers to differentiate knowledge quality attributes for the content, while they perform better for the schema attributes. In other words, asking the reviewers to assess the knowledge quality based on the way it is presented was shown to be less biased that asking them about what they think about the content.

From the knowledge quality perspective, based on a reviewer's bias, it can be concluded that defining knowledge quality based on the content may not be as unambiguous as defining it based on the schema. Despite the recognizable quality metrics that are used for data and information quality, the knowledge content quality is assessed more holistically within a context rather than being observed as mutually exclusive quality attributes. In other words, the schema attributes that refer to different parts of a manuscript can generate more stable quality assessment criteria for a larger number of reviewers.

Finally, from the decision scaling dimension, although there are changes in the decision output made for knowledge quality assessment, nevertheless, the percentage of these changes was not significantly affected by the changes in the sample size. Both content and schema quality evaluations showed that, despite the changes in the sample size, the trends for convergence and divergence pattern, were not be well-established. Meanwhile, prediction on the schema quality evaluation showed more promising results in comparison with the evaluation from the content quality. It should be noted that the decision scaling is very dependent on the sample size, as a very small sample size might not yield a very reliable output when compared with the larger sample sizes.

7.1.3 Qualitative Findings from the Focus Group Study

While we introduced the quality attributes and analysed them in the quantitative survey, in order to better understand the extent of generalizability for the decision paths within a scientific context (Knowledge Management context for this study), we conducted a second experiment with a different paper (as an assessment) and used a different approach (qualitative instead of quantitative). The qualitative approach using focus group interview could independently evaluate the same set of knowledge quality attributes and let the participants build their own preferences and priorities in a tree structure. The manual tree structure provided the opportunity to check the reliability of the decision trees obtained from the quantitative analysis. Choosing a separate paper from the same scientific domain gave a better confidence in terms of generalizability of the quantitative experiment within a scientific domain. By selecting a few experts in the chosen scientific discipline in the qualitative survey of this study, a reliable point of reference, in terms of preferences and priorities, was created so that a better yardstick could be applied in measuring the reliability of the quantitative experiment which exposed diversity in terms of the skills and background of the participants.

7.2 Significance and Contribution of the Study

This research study investigates one of the important topics in the academic world: the quality of scientific publications through defining knowledge quality. While many studies have focused on data and information quality, very few studies have discussed the role of knowledge quality and the way knowledge is generated and assessed in the scientific world. In addition, the peer-review assessment has been mostly observed from the psychological perspective of biases that originate from the reviewer's characteristics rather than misperceptions in understanding knowledge quality assessment biases.

This study also showed that many of the quality attributes used as knowledge quality attributes to assess a research work are redundant and highly inter-related. In fact, adding more to the quality attributes may not contribute to any better assessment, unless those quality attributes can evaluate a significantly different aspect of knowledge quality. This leads to the importance of the notion that knowledge quality attributes may need be crafted properly for each context in order to fully enhance the knowledge quality assessment process. As in the case of peer-review, the schema quality factors are better predictors at, different scales, for a manuscript when compared to content quality attributes.

Finally, that is always the notion of subjectivity in knowledge quality assessment, where every person may think very differently and no agreement can be made on quality of knowledge within a context. This is usually known as low inter-rater validity as earlier mentioned in the literature review and during the analysis. This study showed that the inter-rater validity is subject other elements, such as the background of the reviewer and how closely they are related to a subject. In other words, the selected reviewer population, at any scale, can have different levels of agreement, where the predictability of a decision from reviewers at the core of a scientific discipline can generate more consistent decisions in comparison with those reviewers being related or outside of a knowledge domain. Although the small dataset in this study may not provide a very large scale comparison for generalization, nevertheless, by focusing on a specific knowledge domain, Knowledge Management, it can provide a knowledge quality assessment approach for experts in order to develop better mechanisms on the predictability and evaluation of peer-review decisions.

7.3 Limitations

There were some limitations in both conducting the survey and understanding analysis in this research. To conduct the survey, in order to have a point of reference for comparison, we asked all the participants to read an article and reflect on their evaluation. That made the survey to take a longer time to complete than in a normal survey and this could lead to the response rate being lower than expected. Moreover, this study has only reflected on the knowledge quality measures in a quantified format which may be better explained by mining open ended questions using sentiment analysis for better understanding in the levels of correlation between different quality attributes.

On the other hand, one of the drawbacks in decision trees for small datasets is the generalizability of the decision structure due to what is known as "Overfitting". Overfitting occurs when the decision tree looks for its maximum accuracy by even finding a single case sufficient for branching. In large datasets, in order to control the size of the tree, there is a requirement for specifying the minimum and maximum number of nodes for the parent and child nodes. This means that the decision tree won't branch out unless it can accumulate cases for each node in order to pass the threshold for related parents and child nodes. The bigger the threshold number for the parent and child nodes, the less chance for overfitting, and hence a higher chance of generalizability of the tree structure for similar attributes; Because, the number of observations for a pattern has been repeated several times for each path to give enough confidence on the survivability of the decision path in other future cases. In this study, for some decisions like accept or reject, there were only very few cases (totally 4 out of 27), and these thresholds were set to a minimum of one case occurrence for every node. Although a growth method like CRT can provide a 100% coverage of the cases, nevertheless, a bigger dataset is required to assure the reliability of the long term existence of decisions paths which come into existence due to small threshold requirements. A bigger sample with more diversity and frequency of occurrence in all decision outcomes can reduce this effect.

7.4 Future Research and Recommendations

This study has provided a small workable prototype for evaluating the knowledge quality attributes within a context. There are mainly two areas that can be expanded for future studies, based on what has been analysed and discussed in this research.

The first potential area to expand can be in the design of the questionnaire by having open-ended questions alongside the quantifiable measurement (usually close-ended questions). Such adoption will need to deploy a mechanism that can convert non-quantifiable measures into quantifiable formats. Using text mining and applying sentiment analysis techniques can help in identifying the degree of agreement or disagreement independently from the grading scale method which was used as the only mean of assessment in this study. Such adoption may reveal some hidden aspects of biases in what a reviewer believes and how he/she reflects that belief. This can also help to understand how close or different is the grading of knowledge quality attributes and reviewer's understanding from those knowledge quality attributes.

The second area is on using very large datasets with multitudes of attributes for knowledge quality. In fact, by having a wide range of background information and a lot of knowledge quality attributes, an appropriate method can identify and extract the knowledge quality attributes that best fit within a context. This can be the individual level of personal learning practices in which a person chooses to acquire knowledge through learning, or it can be at a group level of agreement and disagreement in making a decision similar to the peer-review scenarios. It can also be at larger levels for strategic planning and policy design through understanding priorities the organizational or national levels. The big data approach to knowledge quality may be a promising approach in identifying and prioritising the best fitting knowledge quality attributes.

APPENDIX A: Journal Ranking for Knowledge Management

Rank	Title	Citation Score	2008 rank
1	Journal of Knowledge Management	1,284	1
2	Knowledge Management Research & Practice	962	3
3	International Journal of Knowledge Management	880	4
4	Journal of Intellectual Capital	846	2
5	Journal of Information and Knowledge Management	769	9
6	The Learning Organization	717	5
7	Journal of Knowledge Management Practice	651	7
8	Knowledge and Process Management: The Journal of Corporate Transformation	625	6
9	International Journal of Learning and Intellectual Capital	578	10
10	Electronic Journal of Knowledge Management	573	8
11	VINE: The Journal of Information and Knowledge Management Systems	568	14
12	International Journal of Knowledge and Learning	503	12
13	International Journal of Knowledge Management Studies	497	11
14	International Journal of Knowledge, Culture and Change Management	460	13
15	International Journal of Knowledge-Based Development	447	N/A
16	International Journal of Knowledge-Based Organizations	443	N/A
17	Interdisciplinary Journal of Information, Knowledge and Management	424	N/A
18	Knowledge Management & E-Learning: An International Journal	411	N/A
19	Knowledge Management for Development Journal	390	17
20	International Journal of Knowledge Society Research	359	N/A
21	Open Journal of Knowledge Management	349	N/A
22	International Journal of Knowledge and Systems Science	338	N/A
23	actKM: Online Journal of Knowledge Management	329	N/A
24	The IUP Journal of Knowledge Management (formerly The ICFAI Journal of Knowledge Management)	328	18
25	Intangible Capital	304	N/A

APPENDIX B: Quantitative Survey Questions

Q1 Please select your academic position

- ❍ Professor
- ❍ Associate Professor
- ❍ Assistant Professor
- ❍ Senior Lecturer
- ❍ Lecturer
- ❍ Teaching Fellow
- ❍ Others (Please specify) ____________________

Q2 Your Geographical Location

- ❍ Asia
- ❍ Africa
- ❍ Australia and Oceania
- ❍ Europe
- ❍ North America
- ❍ Latin America and Caribbean

Q3 Have you ever been involved in the peer review process for any research work for any scientific journal?

- ❍ Yes
- ❍ No

Q4 Are you an editorial member of any journal?

- ❍ Yes
- ❍ No

Q5 Have you been or a regular reviewer or an editorial member for any of the following journals in the past 5 years? If yes, please select those journals. (Please proceed without choosing any if you have never been an editorial member or a frequent reviewer of any of journal in the following list.

- ❑ actKM: Online Journal of Knowledge Management
- ❑ International Journal of Knowledge and Systems Science
- ❑ Corporate Transformation
- ❑ Electronic Journal of Knowledge Management
- ❑ Intangible Capital
- ❑ Interdisciplinary Journal of Information, Knowledge and Management
- ❑ International Journal of Knowledge and Learning
- ❑ International Journal of Knowledge and Systems Science
- ❑ International Journal of Knowledge Management
- ❑ International Journal of Knowledge Management Studies
- ❑ International Journal of Knowledge Society Research
- ❑ International Journal of Knowledge, Culture and Change Management
- ❑ International Journal of Knowledge-Based Development
- ❑ International Journal of Knowledge-Based Organizations
- ❑ International Journal of Learning and Intellectual Capital
- ❑ Journal of Information and Knowledge Management
- ❑ Journal of Intellectual Capital
- ❑ Journal of Knowledge Management
- ❑ Journal of Knowledge Management Practice
- ❑ Knowledge and Process Management: The Journal of
- ❑ Knowledge Management & E-Learning: An International Journal
- ❑ Knowledge Management for Development Journal
- ❑ Knowledge Management Research & Practice
- ❑ Management Systems
- ❑ Open Journal of Knowledge Management
- ❑ The Learning Organization
- ❑ VINE: The Journal of Information and Knowledge
- ❑ Data & Knowledge Engineering
- ❑ Data Mining and Knowledge Discovery
- ❑ Expert Systems: The Journal of Knowledge Engineering
- ❑ IEEE Transactions on Knowledge and Data Engineering
- ❑ Information, Knowledge, Systems Management
- ❑ International Journal of Applied Knowledge Management
- ❑ International Journal of Human Capital and Information Technology Professionals

- ❑ International Journal of Information Technology and Knowledge Management
- ❑ International Journal of Knowledge-Based and Intelligent Engineering Systems
- ❑ International Journal of Nuclear Knowledge Management
- ❑ International Journal of Software Engineering and Knowledge Engineering
- ❑ International Journal of Technology, Knowledge and Society
- ❑ Journal of Data Mining and Knowledge Discovery
- ❑ Journal of e-Learning and Knowledge Society
- ❑ Journal of Human Capital
- ❑ Journal of Human Resource Costing & Accounting
- ❑ Journal of Knowledge Management, Economics and Information Technology
- ❑ Journal of Knowledge-Based Innovation in China
- ❑ Journal of Universal Knowledge Management
- ❑ Knowledge and Information Systems: An International Journal
- ❑ Knowledge and Innovation: Journal of the KMCI
- ❑ Knowledge, Technology & Policy
- ❑ Knowledge-Based Systems
- ❑ Management Learning: The Journal for Managerial and Organizational Learning
- ❑ Social Epistemology: A Journal of Knowledge, Culture and Policy
- ❑ The Knowledge Engineering Review

Q6 Select your editorial position for each of the previously mentioned journals (Drag and drop the journal name to the appropriate category)

Q7 Type the full name (NOT Abbreviation or short name) of up to 5 journals (if any) that you have been involved in their peer review and was not listed in the previous question. (Skip this question by clicking Next if there is no other journal to be mentioned)

First Journal

Second Journal

Third Journal

Forth Journal

Fifth Journal

Q8 Select your editorial position for each of the previously mentioned journals (Drag and drop the journal name to the appropriate category)

Q9 Approximately how many years you are involved in scientific peer review process ?

- ❍ Less than 3 years
- ❍ 3 to 5 years
- ❍ 5 to 10 years
- ❍ More than 10 years

Q10 Choose one or more from the following options which are closely related to you areas of your expertise.

- ❑ Knowledge Management
- ❑ Artificial Intelligence
- ❑ Information Technology
- ❑ Data Mining/ Predictive Analytics
- ❑ Semantic Web and Technologies
- ❑ Computer Science/Software Engineering
- ❑ Business and Management
- ❑ Intellectual Capital
- ❑ Accounting and Finance
- ❑ Economics
- ❑ Intellectual Property
- ❑ System Thinking and Modelling
- ❑ Others

Please click on the following link. A new browser window will open. The sample article is in PDF format. Read it carefully as if you are a reviewer for this article. Once http://www.emeraldinsight.com/doi/full/10.1108/13673271311315196

Q11 Based on the article, please grade each of the following criteria according to your understanding. (Drag and move the bar between 0 to 100 for each item)

______ Originality: Does the paper contain new and significant information adequate to justify publication?

______ Relationship to Literature: Does the paper demonstrate an adequate understanding of the relevant literature in the field and cite an appropriate range of literature sources? Is any significant work ignored?

______ Methodology: Is the paper's argument built on an appropriate base of theory, concepts, or other ideas? Has the research or equivalent intellectual work on which the paper is based been well designed? Are the methods employed appropriate?

______ Results: Are results presented clearly and analysed appropriately? Do the conclusions adequately tie together the other elements of the paper?

______ Implications for research, practice and/or society Does the paper identify clearly between any implications for research, practice and/or society? Does the paper bridge the gap between theory and practice? How can the research be used in practice (economic and commercial impact), in teaching, to influence public policy, in research (contributing to the body of knowledge)? What is the impact upon society (influencing public attitudes, affecting quality of life)? Are these implications consistent with the findings and conclusions of the paper?

______ Quality of Communication & Language: Does the paper clearly express its case, measured against the technical language of the field and the expected knowledge of the journal's readership? Has attention been paid to the clarity of expression and readability, such as sentence structure, jargon use, acronyms, etc.

Q12 Based on the article, please grade each of the following criteria according to your understanding. (Drag and move the bar between 0 to 100 for each item)

______ Popularity of the subject

______ Appropriateness for the journal

______ Adequacy of literature review

______ Quality of research design

______ Adequacy of data analysis

______ Contributions to the literature

______ Legitimacy of conclusions

______ Practical/managerial significance

______ Clarity of presentation

Q13 Based on the article, please grade each of the following criteria according to the structure of each section of the manuscript. (Drag and move the bar between 0 to 100 for each item)

______ Abstract

______ Introduction

______ Methodology

______ Results

______ Conclusion and Discussion

______ Adequate References

______ Accurate References

______ Up-to-date References

Q14 What would be your overall grade to this manuscript ? (Drag and move the bar between 0 to 100 for each item)

______ Overall Grade

Q15 What is your decision for this manuscript?

- ❍ Accept
- ❍ Minor Revision
- ❍ Major Revision
- ❍ Reject

Q16 Please rank the items below based on the priorities you take into consideration for making a decision about a manuscript. (Drag and drop each item on the location of your preferred ranking)

______ Advancement of knowledge

______ Novelty in the proposed ideas

______ Validity of the proposed methodology

______ Reliability of the proposed hypothesis and research framework

______ Generalizability of the experiment

______ Applicability of the research topic

Q17 Please rank the items below based on the priorities you take into consideration for making a decision about a manuscript. (Drag and drop each item on the location of your preferred ranking)

______ Abstract

______ Introduction and Literature Review

______ Research Framework and Methodology

______ Discussion and Conclusion

______ References

If you wish to be informed of the outcome of this research, you may provide your name and email below, else, you may leave it blank and click "Next"

First Name

Last Name

Email

APPENDIX C: Qualitative Survey Questions and Decision Tree Sample

A. *Assessment Form*

Grade between 0 to 100

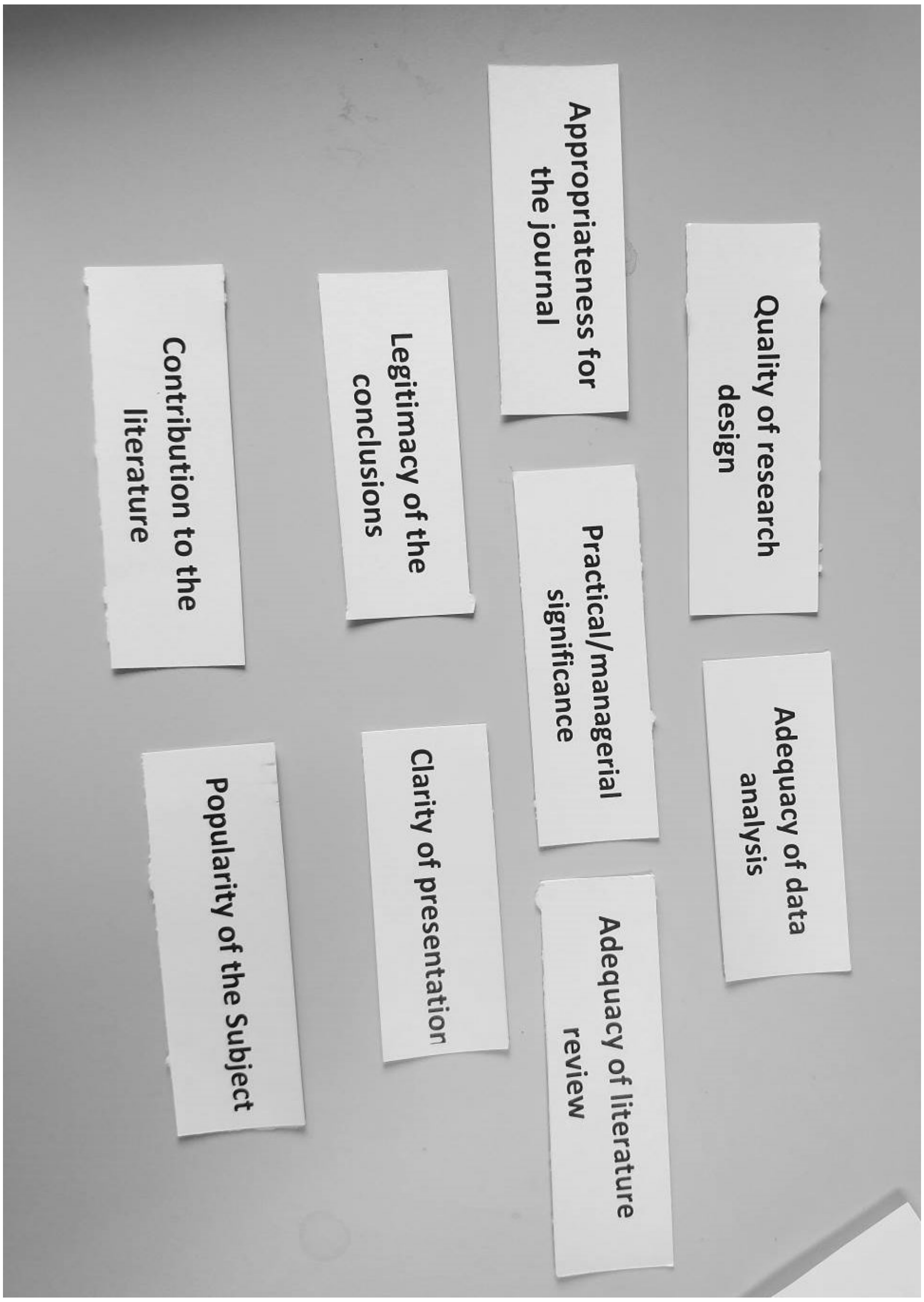
Originality	
Relationship to Literature	
Methodology	
Result	
Implications for research, practice and/or society	
Quality of Communication & Language	

Popularity of the Subject	
Quality of research design	
Legitimacy of the conclusions	
Appropriateness for the journal	
Adequacy of data analysis	
Practical/managerial significance	
Adequacy of literature review	
Contribution to the literature	
Clarity of presentation	

Abstract	
Introduction	
Methodology	
Results	
Conclusion and Discussion	
Adequate References	
Accurate References	
Up-to-date References	

B. Manual Decision Tree Building Using Cards

Low Priority ← **High Priority**



References

Aggestam, L., Backlund, P., & Persson, A. (2010). Supporting Knowledge Evaluation to Increase Quality in Electronic Knowledge Repositories. *International Journal of Knowledge Management*, *6*(1), 23–43.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly*, *25*(1), 107–136. Retrieved from http://www.jstor.org/stable/3250961

Arkoudas, K., & Bringsjord, S. (2009). Propositional attitudes and causation. *International Journal of Software and Informatics*, *3*(1), 47–65.

Armbruster, C. (2008). Open Access in the Natural and Social Sciences: the correspondence of innovative moves to enhance access, inclusion and impact in scholarly communication. *Policy Futures in Education*, *6*(4), 424. doi:10.2304/pfie.2008.6.4.424

Audi, R. (2011). *Epistomology: A Contemporary Introduction to the Theory of Knowledge* (3rd ed.). New Tork: Taylor & Francis.

Bardy, A. H. (1998). Bias in reporting clinical trials. *British Journal of Clinical Pharmacology*, *46*(2), 147–50. doi:10.1046/j.1365-2125.1998.00759.x

Beckman, T. (1998). Knowledge management seminar notes. *ITESM, Monterrey*.

Benda, W. G. G., & Engels, T. C. E. (2010). The predictive validity of peer review : A selective review of the judgmental forecasting qualities of peers , and implications for innovation in science. *International Journal of Forecasting*, *27*(1), 166–182. doi:10.1016/j.ijforecast.2010.03.003

Benos, D. J., Bashari, E., Chaves, J. M., Gaggar, A., Kapoor, N., LaFrance, M., …

Zotov, A. (2007). The ups and downs of peer review. *Advances in Physiology Education*, *31*(2), 145–52. doi:10.1152/advan.00104.2006

Biagioli, M. (1998). The instability of authorship: credit and responsibility in contemporary biomedicine. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *12*(1), 3–16.

Black, T. R. (1999). *Doing Quantitative Research in the Social Sciences: An Integrated Approach to Research Design, Measurement and Statistics*. SAGE Publications Ltd.

Blackburn, J. L., & Hakel, M. D. (2006). An examination of sources of peer-review bias. *Psychological Science*, *17*(5), 378–82. doi:10.1111/j.1467-9280.2006.01715.x

Blank, R. M. (1991). The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review. *The American Economic Review*, *81*(5), 1041–1067. doi:10.2307/2006906

Bohn, R. E. (1998). Measuring and Managing Technological Knowledge. In *The Economic Impact of Knowledge* (pp. 295–314). Butterworth-Heinemann.

Bornmann, L. (2008). Scientific Peer Review : An Analysis of the Peer Review Process from the Perspective of Sociology of Science Theories. *HUMAN ARCHITECTURE: JOURNAL OF THE SOCIOLOGY OF SELF-KNOWLEDGE*, *6*(2), 23–38.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, *45*(1), 197–245. doi:10.1002/aris.2011.1440450112

Bornmann, L., & Daniel, H.-D. (2009). The luck of the referee draw: the effect of exchanging reviews. *Learned Publishing*, *22*(2), 117–125. doi:10.1087/2009207

Bornmann, L., & Daniel, H.-D. (2010). Do author-suggested reviewers rate submissions more favorably than editor-suggested reviewers? A study on atmospheric chemistry and physics. *PLoS ONE*, *5*(10), e13345. doi:10.1371/journal.pone.0013345

Bornmann, L., & Mungra, P. (2011). Improving peer review in scholarly journals. *European Science Editing*, *37*(2), 41–43. doi:10.4103/0256-4602.60162.3

Braben, D. . (2004). *Pioneering research: A risk worth taking*. NJ: Wiley-Interscience.

Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, *23*(1), 4–6. doi:10.1016/j.tree.2007.07.008

Burton-Jones, A., Storey, V. C., Sugumaran, V., & Ahluwalia, P. (2005). A Semiotic Metrics Suite for Assessing the Quality of Ontologies. *Data & Knowledge Engineering*, *55*(1), 84–102.

Cokol, M., Ozbay, F., & Rodriguez-Esteban, R. (2008). Retraction rates are on the rise. *EMBO Reports*, *9*(1), 2. doi:10.1038/sj.embor.7401143

Courtney, J. F. (2001). Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS. *Decision Support Systems*, *31*(1), 17–38.

Davenport, B. T. H., & Prusak, L. (2000). *Working Knowledge: How Organizations Manage What They Know*. Harvard Business Review Press.

Dey, A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing*, *5*(1), 4–7. doi:10.1007/s007790170019

Dietrich, F., & List, C. (2010). The aggregation of propositional attitudes : towards a general theory. In *Oxford Studies in epistomology* (Vol. 3, pp. 215–234).

English, L. P. (2009). *Information Quality Applied: Best Practices for Improving Business Information, Processes and Systems* (Vol. 1st). Indiana: Wiley.

Eppler, M. J. (2006). *Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes* (Vol. 2nd). Berlin: Springer.

Ettenson, R., & Shanteau, J. (1987). Expert Judgment: Is More Information Better? *Psychological Reports*, *60*, 227–238. doi:10.2466/pr0.1987.60.1.227

Evans, J., & Lindsay, W. (2005). *The management and control of quality* (6th ed.). Manson, OH: Thomson.

Ford, E. (2013). Defining and Characterizing Open Peer Review: A Review of the Literature. *Journal of Scholarly Publishing*, *44*(4), 311–326. doi:10.3138/jsp.44-4-001

Garvin, D. (1996). Competing on eight dimensions of quality. *IEEE Engineering Management Review*, *24*(1), 15–23.

Garvin, D. . (1984). What does product quality really mean? *Sloan Management Review*, *26*(1), 25–43.

Godlee, F. (2000). The ethics of peer review. In *Ethical Issues in Biomedical Publication* (1st ed., pp. 59–84). Johns Hopkins University Press.

Goodman, S. N., Berlin, J., Fletcher, S. W., & Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Ann Intern Med*, *121*(1), 11–21. doi:10.7326/0003-4819-121-1-199407010-

00003

Gorman, G. E. (1999). Library and information science journals in the Asian context. *IFLA Council and General Conference*. Retrieved from http://www.ifla.org/IV/ifla65/papers/005-118e.htm

Gorman, G. E., & Calvert, P. J. (2003). LIS Journal Quality: Results of a Study for the IFLA Library and Information Science Journals. *IFLA Council and General Conference*, (69). Retrieved from http://www.ifla.org/IV/ifla69/papers/208e-Gorman_Calvert.pdf

Hardaway, D. E., & Scamell, R. W. (2012). Open Knowledge Creation: Bringing Transparency and Inclusiveness to the Peer Review Process. *MIS Quarterly*, *36*(2), 339–346.

Hargens, L. L. (1988). Scholarly Consensus and Journal Rejection Rates. *American Sociological Review*, *53*(1), 139–151. doi:10.2307/2095739

Hargens, L. L. ., & Herting, J. R. . (1990). Neglected considerations in the analysis of agreement among journal referees. *Scientometrics*, *19*(1), 91–106. doi:10.1007/BF02130467

He, W., & Wei, K.-K. (2009). What drives continued knowledge sharing? An investigation of knowledge-contribution and -seeking beliefs. *IT Decisions in Organizations*, *46*(4), 826–838. doi:10.1016/j.dss.2008.11.007

Hojat, M., Gonnella, J. S., & Caelleigh, A. S. (2003). Impartial Judgment by the "Gatekeepers" of Science: Fallibility and Accountability in the Peer Review Process. *Advances in Health Sciences Education*, *8*(1), 75–96. doi:10.1023/A:1022670432373

Holsapple, C. W., & Joshi, K. D. (2001). Organizational knowledge resources. *Decision Support Systems*, *31*(1), 39–54. doi:10.1016/s0167-9236(00)00118-4

Horrobin, D. F. (1990). The Philosophical Basis of Peer Review and the Suppression of Innovation. *JAMA: The Journal of the American Medical Association*, *263*(10), 1438. doi:10.1001/jama.1990.03440100162024

How Science Goes Wrong. (2013, October). *Economist*. Retrieved from http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong

Huang, K. T., Lee, Y. W., & Wang, R. Y. (1999). Create Organizational Knowledge. In *Quality Information and Knowledge* (pp. 91–110). New Jersey: Prentice Hall.

Jackson, J. L., Srinivasan, M., Rea, J., Fletcher, K. E., & Kravitz, R. L. (2011). The validity of peer review in a general medicine journal. *PLoS ONE*, *6*(7), e22475. doi:10.1371/journal.pone.0022475

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*(1), 1–12. doi:10.1016/0197-2456(95)00134-4

Jarke, M., Jeusfeld, M., Quix, C., & Vassiliadis, P. (1999). Architecture and quality in data warehouses: An extended repository approach. *Information Systems*, *24*(3), 229–253.

Jefferson, T., Rudin, M., Brodney Folse, S., & Davidoff, F. (2007). Editorial peer review for improving the quality of reports of biomedical studies. *The Cochrane Database of Systematic Reviews*, (2), MR000016. doi:10.1002/14651858.MR000016.pub3

Jefferson, T., Wager, E., & Davidoff, F. (2002). Measuring the quality of editorial peer review. *JAMA : The Journal of the American Medical Association*, *287*(21), 2786–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12038912

Juran, J. (1992). *Juran on quality by design*. New York: The Free Press.

Juran, J. (1992). *Juran on quality by design: the new steps for planning quality into goods and services*. Simon and Schuster.

Kothari, C. (2009). *Research Methodology* (2nd ed.). New Age International Pvt Ltd Publishers.

Kotu, V., & Deshpande, B. (2015). Comparison of Data Mining Algorithms. In *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner* (pp. 407–416). Elsevier.

Labbé, C., & Labbé, D. (2012). Duplicate and fake publications in the scientific literature: how many SCIgen papers in computer science? *Scientometrics*, *94*(1), 379–396. doi:10.1007/s11192-012-0781-y

Lam, A. (2000). Tacit Knowledge, Organizational Learning and Societal Institutions: An Integrated Framework. *Organization Studies*, *21*(3), 487–513. doi:10.1177/0170840600213001

Lamont, M., & Huutoniemi, K. (2011). Comparing customary rules of fairness: Evaluative practices in various types of peer review panels. In L. Camic, Gross (Ed.), *Social Knowledge in the Making* (pp. 209–232). University of Chicago Press.

Langfeldt, L. (2006). The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation*, *15*(1), 31–41.

Retrieved from http://rev.oxfordjournals.org/content/15/1/31.short

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17. doi:10.1002/asi.22784

Lee, J., Lee, Y., & Ryu, Y. (2007). Information Quality Drivers of KMS. *Proceedings of the 2007 International Conference on Convergence Information Technology*. IEEE Computer Society. doi:10.1109/iccit.2007.242

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & Management*, *40*(2), 133–146. doi:10.1016/S0378-7206(02)00043-5

Lemos, N. (2009). *An Introduction to the Theory of Knowledge*. Cambridge: Cambrdige University Press.

Levis, A. W., Leentjens, A. F. G., Levenson, J. L., Lumley, M. A., & Thombs, B. D. (2015). Comparison of self-citation by peer reviewers in a journal with single-blind peer review versus a journal with open peer review. *Journal of Psychosomatic Research*. doi:10.1016/j.jpsychores.2015.08.004

Liebowitz, J. (1999). *Knowledge Management Handbook*. *International handbooks on information systems*. doi:10.1201/b12285

Lim, E.-P., Vuong, B.-Q., Lauw, H. W., & Sun, A. (2006). Measuring Qualities of Articles Contributed by Online Communities. *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE Computer Society.

Link, A. M. (1998). US and non-US submissions: An analysis of reviewer bias.

Journal of the American Medical Association, *280*(3), 246–247. doi:10.1001/jama.280.3.246

Lyons, J. C. (2009). Perception and Virtue Reliabilism. *Acta Analytica*, *24*(4), 249–261. doi:10.1007/s12136-009-0064-2

Mancilla-Amaya, L., Sanín, C., & Szczerbicki, E. (2012). Quality Assessment of Experiential Knowledge. *Cybernetics and Systems*, *43*(2), 96–113. doi:10.1080/01969722.2012.654071

Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *The American Psychologist*, *63*(3), 160–168. doi:10.1037/0003-066X.63.3.160

Maxim, L., & van der Sluijs, J. P. (2011). Quality in environmental science for policy: Assessing uncertainty as a component of policy analysis. *Environmental Science & Policy*, *14*(4), 482–492. doi:10.1016/j.envsci.2011.01.003

Maxwell, R. (1992). Dimensions of quality revisited: from thought to action. *Quality in Health Care*, *1*(3), 171.

McCormack, N. (2009). Peer Review and Legal Publishing: What Law Librarians Need to Know about Open, Single-Blind, and Double-Blind Reviewing. *Law Library Journal*, *101*(59). Retrieved from http://papers.ssrn.com/abstract=1339227

McNutt, R. A. (1990). The Effects of Blinding on the Quality of Peer Review. *JAMA*, *263*(10), 1371. doi:10.1001/jama.1990.03440100079012

Melkas, H., & Harmaakorpi, V. (2008). Data, information and knowledge in regional innovation networks, Quality considerations and brokerage functions. *European*

Journal of Innovation Management, *11*(1), 103–124.

Memmi, D. (2008). The Social Context of Knowledge (Vol. Illustrate, pp. 189–208). Information Science Reference.

Merton, R. K. (1968). The Matthew Effect in Science. *Science*, *159*(3810), 56–63. doi:10.1126/science.159.3810.56

Millie Kwan, M., & Balasubramanian, P. (2003). KnowledgeScope: managing knowledge in context. *Decision Support Systems*, *35*(4), 467–486. doi:10.1016/s0167-9236(02)00126-4

Newton, D. P. (2010). Quality and peer review of research: an adjudicating role for editors. *Accountability in Research*, *17*(3), 130–45. doi:10.1080/08989621003791945

Nicholas, D., Watkinson, A., Jamali, H. R., Herman, E., Tenopir, C., Volentine, R., … Levine, K. (2015). Peer review: still king in the digital age. *Learned Publishing*, *28*(1), 15–21. doi:10.1087/20150104

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. doi:10.1037/1089-2680.2.2.175

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, *5*(1), 14–37. doi:10.1287/orsc.5.1.14

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company*. New York: Oxford University Press.

Owlia, M. S. (2010). A framework for quality dimensions of knowledge management systems. *Total Quality Management & Business Excellence*, *21*(11), 1215–1228. doi:10.1080/14783363.2010.529351

Parasuraman, A. (2002). Service quality and productivity: A synergistic perspective. *Managing Service Quality*, *12*(1), 6–9. doi:10.1108/096045202104

Pierce, E., Kahn, B., & Melkas, H. (2006). A Comparison of Quality Issues for Data, Information, and Knowledge. In *Emerging Trends and Challenges in Information Technology Management* (pp. 60–63). Hershey: Idea Group.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211. doi:10.1145/505248.506010

Polanyi, M. (1967). The tacit dimension.

Polanyi, M. (1974). *Personal Knowledge: Towards a Post-Critical Philosophy*. University of Chicago Press. doi:10.2307/2092944

Popper, K. P. (1972). *Objective Knowledge*. Oxford: Clarendon Press.

Poston, R. S., & Speier, C. (2005). Effective use of knowledge management systems: a process model of content ratings and credibility indicators. *MIS Quarterly.*, *29*(2), 221–244. Retrieved from http://dl.acm.org/citation.cfm?id=2017254.2017258

Rao, L., & Osei-Bryson, K.-M. (2007). Towards defining dimensions of knowledge systems quality. *Expert Systems with Applications*, *33*(2), 368–378.

Redman, T. C., & Godfrey, A. B. (1996). *Data Quality for the Information Age*. Artech House Publishers.

Reeves, C. A., & Bednar, D. A. (1994). Defining Quality: Alternatives and Implications. *The Academy of Management Review*, *19*(3), 419–445. doi:10.2307/258934

Resnik, D. B., & Elmore, S. A. (2015). Ensuring the Quality, Fairness, and Integrity of Journal Peer Review: A Possible Role of Editors. *Science and Engineering Ethics*. doi:10.1007/s11948-015-9625-5

Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., … Krumholz, H. M. (2006). Effect of blinded peer review on abstract acceptance. *JAMA : The Journal of the American Medical Association*, *295*(14), 1675–1680. doi:10.1016/S0093-3619(08)70551-3

Rothwell, P. M., & Martyn, C. N. (2000). Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain : A Journal of Neurology*, *123 (Pt 9*, 1964–1969. doi:10.1093/brain/123.9.1964

Rowland, F. (2002). The peer-review process. *Learned Publishing*, *15*(4), 247–258. doi:10.1087/095315102760319206

Sandström, U., & Hällsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics*, *74*(2), 175–189. doi:10.1007/s11192-008-0211-3

Schultz, D. M. (2010). Are three heads better than two? How the number of reviewers and editor behavior affect the rejection rate. *Scientometrics*, *84*(2), 277–292. doi:10.1007/s11192-009-0084-0

Seawright, K. W., & Young, S. T. (1996). A Quality Definition Continuum. *Interfaces*, *26*(3), 107–113. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=4478268&site=ehost-live

Serenko, A., & Bontis, N. (2013). Global ranking of knowledge management and intellectual capital academic journals: 2013 update. *Journal of Knowledge Management*, *17*(2), 307–326. doi:10.1108/13673271311315231

Shin, M., Holden, T., & Schmidt, R. A. (2001). From knowledge theory to management practice: towards an integrated approach. *Information Processing & Management*, *37*(2), 335–355. doi:10.1016/s0306-4573(00)00031-5

Smaling, A. (1992). Varieties of methodological intersubjectivity — the relations with qualitative and quantitative research, and with objectivity. *Quality and Quantity*, *26*(2), 169–180. doi:10.1007/BF02273552

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, *99*(4), 178–82. doi:10.1258/jrsm.99.4.178

Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology*, *20*(8), 357–358. doi:10.1016/S0167-7799(02)01985-6

Suárez, A., Bernhard, J., & Dellavalle, R. (2012). Hiding Behind the Curtain: Anonomyous Versus Open Peer Review. In L. Bercovitch & C. Perlis (Eds.), *Dermatoethics SE - 36* (pp. 221–225). Springer London. doi:10.1007/978-1-4471-2191-6_36

Travis, G. D. L., & Collins, H. M. (1991). New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System. *Science, Technology & Human Values*, *16*(3), 322–341. doi:10.1177/016224399101600303

Trouble at the lab. (2013, October). *Economist*. Retrieved from http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble

Van Rooyen, S., Godlee, F., Evans, S., Smith, R., & Black, N. (1999). Effect of blinding and unmasking on the quality of peer review. *Journal of General Internal Medicine*, *14*(10), 622–624. doi:10.1046/j.1525-1497.1999.09058.x

Van Selm, M., & Jankowski, N. W. (2006). Conducting online surveys. *Quality and Quantity*, *40*(3), 435–456. doi:10.1007/s11135-005-8081-8

Ware, M. (2008). Peer review: benefits, perceptions and alternatives. *Publishing Research Consortium*, *4*.

Ware, M. (2011). Peer Review: Recent Experience and Future Directions. *New Review of Information Networking*, *16*(1), 23–53. Retrieved from http://www.tandfonline.com/doi/citedby/10.1080/13614576.2011.566812

Weber, E. J., Katz, P. P., Waeckerle, J. F., & Callaham, M. L. (2002). Author Perception of Peer Review: Impact of Review Quality and Acceptance on Satisfaction. *JAMA: The Journal of the American Medical Association*, *287*(21), 2790–2793. doi:10.1001/jama.287.21.2790

Wenger, E., McDermot, R., & Snyder, W. M. (2002). *Cultivating Communities of Practice*. Boston: Harvard Business Review Press.

Wenger, E., White, N., & Smith, J. D. (2009). *Digital Habitats: Stewarding technology for communities*. Portland, OR: CPsquare.

Wiig, K. M. (1994). Knowledge Management Foundations: Thinking about Thinking - how People and Organizations Represent, Create, and Use Knowledge. Retrieved from http://dl.acm.org/citation.cfm?id=528424

Yoo, D. K., Vonderembse, M., & Ragu-Nathan, T. S. (2010). Knowledge quality: antecedents and consequence in project teams. *Journal of Knowledge Management*, *15*(2), 329–343.

Zalta, E. (2012). The Analysis of Knowledge. In *The Stanford Encyclopedia of Philosophy*.

Zhuang, Z., Elmacioglu, E., Lee, D., & Giles, C. L. (2007). Measuring conference quality by mining program committee characteristics. In *Proceedings of the 2007 conference on Digital libraries - JCDL '07* (p. 225). New York, New York, USA: ACM Press. doi:10.1145/1255175.1255220

Zimmermann, A., Lorenz, A., & Oppermann, R. (2007). An Operational Definition of ContextModeling and Using Context. In B. Kokinov, D. Richardson, T. Roth-Berghofer, & L. Vieu (Eds.), (Vol. 4635, pp. 558–571). Springer Berlin / Heidelberg. doi:10.1007/978-3-540-74255-5_42