

Centre for Geo-Information

Thesis Report GIRS-2007-25

Reviewing the book chapter reviews
of “An Introduction to Spatial Data Infrastructures”

Mark Blomsma

August 2007



WAGENINGEN UNIVERSITY

WAGENINGEN UR



Reviewing the book chapter reviews
of “An Introduction to Spatial Data Infrastructures”

Mark Blomsma

Registration number 83 11 02 076 090

Supervisors:

Dr. Ir. Joep Crompvoets

Prof. Dr. Ir. Arnold Bregt

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

August 2007

Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2007-25
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Foreword

When the idea for this thesis formed in my mind, I had doubts about the usefulness for the GIS department. Therefore I would like to thank Joep Crompvoets and Arnold Bregt for giving me the opportunity to develop my ideas into an actual thesis. Though the link with GIS may seem a bit far fetched, I believe that the thesis can be of assistance in book writing for all fields of science, including GIS. In the creation of this thesis I have become painfully aware of the need for work on peer review in the scientific world. I sincerely hope that this thesis has made a valid contribution to the book 'An Introduction to Spatial Data Infrastructures' and I wish both Joep and Arnold the best of luck getting the book published.

Furthermore I would like to thank my friends and family for their support. They were always more than happy to provide me with some diversion whenever I got tired of working. And last but not least I would like to thank my fiancée for sitting by my side during the hours when it seemed I only had attention for my computer screen.

Table of content

Chapter 1	Introduction	1
Chapter 2	Methodology	4
	<i>The review process: peer bias and quality</i>	5
	<i>Expert knowledge</i>	6
	<i>Review evaluation instruments</i>	7
	<i>Unofficial assessment instruments</i>	10
	<i>Review methodology options</i>	12
	<i>Creating Method 3: Grade sheet for analysing book reviews</i>	13
	<i>Creating method 4: Methodology description for analysing book chapter reviews per comment</i>	15
	<i>Methodology for stepwise analyses of the results</i>	17
	<i>Step 1: Methods 1, 2 and 3 compared to author's grades</i>	17
	<i>Step 2: Separate comments of method 4 and of the best methods of step 1 compared to author's accepted separate comments</i>	18
	<i>Step 3: Final analyses of the methods comparing quality with time consumption</i>	18
Chapter 3	Results & Interpretation	19
	<i>Step 1</i>	19
	<i>Step 1A: Methods 1 through 3 compared to baseline using acceptance/rejection results</i>	19
	<i>Step 1B Methods 2 and 3 compared to baseline(grades) using actual grades</i>	20
	<i>Part 1: Basic attributes</i>	20
	<i>Part 2: Correlation</i>	20
	<i>Part 3: Regression analyses</i>	20

<i>Step 1C: Calibration of method 3</i>	21
<i>Part 1: Methods 1, 2 and calibrated method 3 compared to baseline using acceptance/rejection results</i>	21
<i>Part 2: Basic attributes</i>	22
<i>Part 3: Correlation and regression of calibrated method 3</i>	22
<i>Step 2</i>	23
<i>Step 3: Analysing methods: quality versus time consumed</i>	26
<i>Presenting a summary of the main suggestions for improving the book Chapters</i>	28
Chapter 4 Conclusion	30
<i>Method 1</i>	30
<i>Method 2</i>	30
<i>Method 3</i>	30
<i>Method 4</i>	31
<i>Main objective: To analyse the student reviews in order to present an overview which will help improve the quality of the book 'An introduction to Spatial Data Infrastructures'.</i>	31
<i>Discussion and recommendations</i>	32
List of cited literature	33
Appendix A	35
Appendix B	41
Appendix C	43

List of tables

Table 2.1	Chapter of the book 'An Introduction to Spatial Data Infrastructures'	4
Table 3.1	Percentages of similarity in acceptance of reviews between method 1 through 3 and baseline using two levels of acceptance	19
Table 3.2	Basic attributes of methods 2,3 and the baseline (grades)	20
Table 3.3	Linear correlation between methods 2 and 3 and the baseline (grades)	20
Table 3.4	Percentage of similarity in acceptance of reviews between method 1, method 2, calibrated method 3 and baseline using two levels of acceptance	21
Table 3.5	Basic attributes of methods 2, calibrated method 3 and the baseline (grades)	22
Table 3.6	The total amount of comments per chapter, as well as those chosen by the separate authors and by both authors	24
Table 3.7	The comments per chapter of the reviews chosen by method 2, as well as those chosen by the separate authors and by both authors	24
Table 3.8	The comments per chapter of the reviews chosen by calibrated method 3, as well as those chosen by the separate authors and by both authors	25
Table 3.9	Percentages of comments accepted by methods and approved by both authors compared to total amount of approved comments by both authors	25
Table 3.10	Percentages of comments approved by both authors compared to the total amount of comments accepted by the methods	26
Table 3.11	The total time needed to analyse the chapter reviews with method 2 and 3 in minutes (between brackets the number of reviews per chapter)	26

List of figures

Figure 2.1	Peer review scoring system (Feurer <i>et al.</i> , 1994)	7
Figure 2.2	Review Quality Instrument (RQI) for assessing peer review of manuscripts (Rooyen <i>et al.</i> , 1999)	8
Figure 2.3	Critical Paper Review Grading Sheet (University of Massachusetts Lowell, Dept of Computer Science, 2007)	11
Figure 2.4	Grading sheet for analysing book chapter reviews based on “Peer review scoring system” (Feurer <i>et al.</i> , 1994)	14
Figure 2.5	Decision flowchart for method 4	16
Figure 2.6	Methodology scheme for result analyses	17
Figure 3.1	Methods 2 and 3 plotted against baseline (grades) with added linear trend lines	21
Figure 3.2	Calibrated method 3 plotted against baseline (grades) with added linear trend line	22

Abstract

In the course of 2006 a study book on SDI, "An Introduction to Spatial Data Infrastructures", has been written. Thirty-five students wrote reviews on the different chapters of this book. The thesis aimed to extract the main suggestions for improving the book chapters from these reviews. It also aimed to develop and analyse methods that can be used in future book writing to analyse book chapter reviews. Four methods were eventually selected for analyses. Two of these methods were taken from existing scientific work. It was then tried to improve the second method, a scoring card, resulting in a third method. These three methods all test reviews as a whole. Using logical deduction instead of existing work, a fourth method was created from scratch that regarded not the whole review, but each single comment in each review. Analyses showed that of the four methods, three yielded better than random results. The existing scoring card method was fast and efficient at selecting reviews. The altered version of this scoring card initially performed worse, but was found to need calibration after which it out-performed the existing scoring card. It was therefore concluded that improving an existing scoring card can be efficient only if enough time for calibration exists. The fourth method was found to select good comments, and the selected comments from this method were indeed presented as main suggestions for improvement of the book chapters. It was however concluded that the selected suggestions were a very small selection and improvement of the method should result in a greater selection of comments and suggestions. It was also attempted to combine the scoring card methods with the loose comment method. It was found that these combinations yielded similar results but took 50% more time and were therefore not recommended for use.

Chapter 1: Introduction

The field of GIS has been growing exponentially in the last few decades. New technology has opened possibilities for massive data intake and storage. Spatial Data Infrastructures (SDI's) are created to cope with these large amounts of data. Though books and papers on the subject of SDI have been written, such as "the SDI Cookbook" (Nebert, 2004), "Geospatial Data Infrastructure, concepts, cases and good practice" (Groot and McLaughlin, 2000) and "Geodaten-Infrastruktur Grundlagen und Anwendungen" (Bernard *et al.*, 2005), study material for future GIS students is still lacking. Prof. Dr. Ir. Arnold Bregt and Dr. Ir. Joep Crompvoets of Wageningen University & Research (WUR) have taken it upon themselves to create such a work. The work, "An introduction to Spatial Data Infrastructures" (2006), encompasses all major issues of SDI such as data, technology, people and policies. First drafts have been made on the various chapters, and a step of refinement has been reached. Thirty-five students of the master course SDI (2006) of WUR have reviewed these chapters, providing input for improvement and extension of each chapter.

This thesis will critically analyse these reviews on objectivity and content and eventually summarize the given suggestions. For this purpose several methodologies for analysing book chapter reviews will be chosen and created and tested on the various chapters. The methods are then analysed and the best methods are used to create the suggestions for the book chapters. The thesis therefore has a duality in final conclusions. The thesis not only aims to develop and analyse methods that can be used in future book writing to analyse book chapter reviews. It also aims to directly improve the book "An introduction to Spatial Data Infrastructures" (2006) by providing the authors an overview of the best suggestions the student reviews have to offer.

The book has 7 chapters. The last chapter provides case studies of different SDI's and will not be used. The first six chapters are on SDI, Data, People, Standards, Technologies and Policies. The suggestions are organized in the same way as the reviews, per chapter.

The history of SDI started when the availability of digital data increased and the need for this data to be accessible for decision makers increased as well. Spatial Data Infrastructures were developed to answer that need. In the US the Executive Order 12906 was signed by Bill Clinton in 1994, effectively giving the start signal for a first National SDI. In the following years, many countries or regions, including the EU (called the INSPIRE initiative) would follow (Crompvoets and Bregt, 2006). In 2002 there were 120 countries that had at least started an initiative towards a national SDI (Crompvoets and Bregt, 2003). Though there is no single definition of an SDI (Williamson *et al.*, 2003), most encompass at least the following: datasets, standards, technology, policies and people (Bregt *et al.*, 2006). Each of these is necessary for the constant evolution that SDI's are still going through every day and therefore each of these concepts is treated in the book in a different chapter. Datasets are needed as they are the actual data the infrastructure is built for. In the case of technology, and especially technological advancements, the link with SDI is also easily recognised. Since digital data is involved, the development of databases and geo-software have given SDI's the possibility to grow. Because of fast and reliable internet, clearinghouses can now be made. These websites are a portal to SDI's all over the world making the data available to many new users and even providing extra services for these users (Crompvoets, 2006). For standards, policies and people it is

harder to see the link with SDI, yet they are essential for any SDI initiative to be successful. Standards ensure the smooth transfer of digital data from one owner to another and allow for interpretation of all spatial data for those who work with the same standards. Policies provide a legal framework for SDI initiatives. Digital data is a new form of possession, and owners need to be ensured of their legal rights before they willingly share their data with others. People are those who will actually use the SDI, they decide what data is needed, they set-up the policies and invent the technology. All these parts together form the SDI that can be used as a powerful tool in decision making. The use of SDI can make working with geo-information more efficient and of higher quality (Bernard *et al.*, 2005). It can also reduce time and costs for the decision making process which is for instance recognised by the EU when making environmental assessments (Vanderhaegen and Muro, 2004). Indeed by providing the data the information is easier to reach and as such more valuable to those who now have access to it (Loenen, 2006).

As the field of SDI has been developing for the last decade, researchers, authors and organisations have written articles and books on SDI and subjects related to this new phenomenon. One such organization is the Global Spatial Data Infrastructure which spreads the so called “SDI Cookbook” though the internet, providing a world wide available and regularly updated manuscript on the development of SDI’s. (Nebert, D., 2004) Other books, such as “Developing Spatial Data Infrastructures” by Williamson *et al.* (2003), provide an overview on recent SDI developments and visions on the near future. Though general works on the subject exist such as “Geospatial Data Infrastructure, concepts, cases and good practice” (Groot and McLaughlin, 2000) and “Geodaten-Infrastruktur Grundlagen und Anwendungen” (Bernard *et al.*, 2005), a study book focussing on future geo-information master students to learn about the subject is still missing. Prof. Dr. Ir. A. Bregt and Dr. Ir. Joep Crompvoets of Wageningen University & Research (WUR) are writing this book, closing this gap in knowledge for the future.

This thesis will contribute to this book, not by submitting new text, but by analysing the reviews done on the books rough draft, hopefully providing a better overview of the given comments and suggestions. The used reviews were made by GIS students from the same university who just completed their SDI course. The students come from varying international backgrounds such as The Netherlands, China, Indonesia, USA and other countries from Europe, Asia, South America and Africa, totalling 35 students. The review format was taken from an existing GIS magazine.

To analyse the quality of the reviews it is necessary to have a well thought through methodology for reviewing the chapter reviews. Reviewing is done differently by people from different nations and disciplines (Joyce *et al.*, 1998), and the students who reviewed the book chapters were culturally mixed indeed. Furthermore ethics play a strong role in reviewing, reviewers must be fair yet firm in their reviews. (Baldwin, 1998) So to avoid complications in analysing these reviews a general method, uninfluenced by culture, is needed to analyse these reviews. To acquire this method several existing methods are sought and tested before applying a single template on all reviews. New methods are created as well. Each method is then tested on the book chapter reviews and analysed through comparison. The comparison input for this analyses are the opinions of the original book authors, they

are seen as the experts of their own book and provide the baseline with which to test the reviews.

The basis for methods can differ, depending on the criteria of the creator and the chosen review template. Generally speaking some of the addressed issues in each methodology are likely to be similar, yet differences can be expected. A research by Hartley (2005) on academic book reviews questions the writers of academic reviews on “what features they thought were important in book reviews”. Highly valued by the writers was a straightforward overview of what the book is about followed by information on intended audience, a substantial discussion and academic credibility among others. This shows the variety of issues that can be addressed in a review assessment methodology.

The academic world has not yet reached a general agreement on how to measure the quality of editorial peer review. Jefferson *et al.* (2002) have investigated and compared several studies on the quality of peer review, only to conclude that though general indicators of quality can be found in most researches, no actual consensus is reached. And that in order to improve quality of research, efficient systems for evaluating scientific information of peer reviews must be created.

The objective of this thesis is to analyse the student reviews in order to present an overview which will help improve the quality of the book “An introduction to Spatial Data Infrastructures”. The following sub-objectives can be distinguished.

1. Collecting the review results
2. Finding several methodologies for reviewing book/paper reviews
3. Creating own method(s)
4. Analysing all chapters with found/created methods
5. Analysing methods: quality versus time consumed
6. Presenting a summary of the main suggestions for improving the book chapters
7. Presenting conclusions and suggestions for improvement on the methodologies

The research questions that relate to these objectives are:

- What methodologies exist for reviewing book reviews?
- What other method(s) can be thought off?
- Which methodology fits the used review approach the best?
- What are the main suggestions given in the chapter/case reviews?

Even though a consensus on editorial peer review assessment has not been reached in the academic world, attempts to quantify the quality of such reviews have been done. In order to analyse the book chapter reviews of this thesis, such fixed methods must be used. The fact that no consensus exists (Jefferson *et al.*, 2002) poses a problem as there is no immediate way to evaluate the assessment criteria and therefore the quality of the methods. In order to quantify the quality of each method, the outcome of the methods is compared to the choices the actual authors would make when choosing reviews to use in the process of improving their book. Though this seems rather subjective, it is the most direct way of evaluation as long as the academic world has not created a fixed set of rules for peer review.

Chapter 2: Methodology

Input for the analyses are the reviews provided by the 2006 master course of SDI from Wageningen University. 35 students from different nationalities provided reviews on the different chapters of the book 'An introduction to Spatial Data Infrastructures'. The following table shows the amount of reviews per chapter.

Chapter number	Chapter name	Number of reviews
1	What is SDI?	4
2	(Spatial) Data	6
3	People	4
4	Standards	6
5	Technology	5
6	Policies	4
7	Cases	6

Table 2.1: Chapters of the book 'An introduction to Spatial Data Infrastructures'

The reviews were strongly in line with an existing format of the International Journal of Geographic Information Science. (for reviews see Appendix A) The reviews consist of 3 parts: 1. General questions; 2. Grading of content and 3. Detailed comments and suggestions. The first part has nine questions on the quality of: Grammar, Writing style, Titles, Introduction, Chapter Length. Conclusions, Tables&Figures, References and Questions&Exercises. Students have selected an answer and provided some small comments. The second part required the students to grade the following issues from 1 (bad) to 10 (perfect): originality, didactic level, technical quality, relevance to SDI, presentation and overall rating and lastly asked for personal confidence level on these ratings (low, average, high). In the last part students have provided an approximately 2 page review of the chapter providing detailed comments on both textual and contextual issues giving suggestions for changes, errors and possible extensions.

The last chapter (7), Cases, did not receive regular reviews. Instead students wrote reviews of the SDI's of their respective countries. In order to use only reviews of the same format, these reviews are left out of the analyses.

Journals often rely on peer review to evaluate articles. In case of Medical Journals correct reviewing can mean the difference between life and death. It is therefore no surprise that much of the literature on peer reviewing has been written by Medical specialists and authors. Bias and quality of this process and the time involved are important issues when discussing peer review. (Pierie *et al.*, 1996) The reason for this is that although there are fixed set of rules available for authors, there are no such rules for the peer review process to ensure its quality.

Creating such a method to evaluate peer reviews, in this case book chapter reviews, is the aim of this research. In describing the methodology for creating such a method first the elimination of peer bias in this review process will be evaluated using existing literature. The process of finding a methodology for evaluating reviews must be as unbiased as possible, only then can the quality of reviews be measured. Second, expert knowledge will be used in order to provide a guideline for creating a review evaluation method. Then existing methods are evaluated. From those methods some methods will be chosen to be tested or adapted. And finally one method will be described that has never been described in existing literature.

The review process: peer bias and quality

In the instance of this thesis, two forms of peer bias can be distinguished. One of the reviewers that have reviewed the book chapters. The other of the evaluator that will evaluate the reviews. Both are acquainted with the authors of the work they are reviewing/analysing and are therefore likely to be biased. The following list of researches will provide ways to minimize bias in the evaluation process. For each research the conclusions will be described followed by the effect this has on the chosen methodology of this thesis.

Fisher *et al.* (1994) found that when peers are not aware who the author of an article is (blinded peer review) they are more likely to provide an unbiased peer review. Though this thesis analyses reviews and does not review actual articles, it can be assumed that a blinded analyses of the reviews will decrease any possible bias and therefore improve the quality of the thesis. Names will therefore be removed from the reviews in order to minimize bias.

Not all bias however will be removed as the same study (Fisher *et al.* (1994)) also (in accordance with previous studies) showed that half of the “blinded” reviewers were still able to identify the authors. They could identify the author through subject or writing style. The reviews for this thesis are all on the same subject and therefore subject will not provide a clue on the identity of the author. Writing style could provide some clue on recognition as the reviewers and the evaluator have been colleagues. However, the evaluator has not seen earlier work of the reviewers, so any style recognition will be from vocal experience. This should again reduce recognition probability, and therefore bias.

Joyce *et al.* (1998) concluded that writing style is also likely to differ due to nationality and background of the reviewers. The population used for reviews was multi-cultural and at least for use of references it has been found that both nationality and background have an influence. This will therefore increase the chance of recognition by the evaluator and as such also increase possible bias.

The reviewers themselves have not been blinded. The reviews were made as an assignment for a course and the reviewers were well aware of the authors of the book. As those same authors graded the assignment, bias can be expected especially for the part where the reviewers had to grade the work of the authors. These grades are likely to be higher then they would have been in a blinded review. Of more interest to this thesis are the detailed comments. Bias in this case will most likely result in constructive comments as useful comments would results in better assignment grades and will therefore provide a less biased part of the reviews.

Blinding of reviews can work in two ways. Not only can the reviewers be blinded on the author of the reviewed work, the authors can also be blinded from the names of the reviewers. This is a common practice in peer reviews. Walsh *et al.* (2000) has shown that blinding of reviewers names for the author of the reviewed work can have adverse effects on the quality of the reviews. In the study those reviewers who had to sign their reviews spend more time on their reviews, were more courteous in the review, had a significantly better review quality and were more likely to accept the reviewed work. It is however discussed that possible improvements might have come only from the fact that the reviewers knew they were participating in

a research. Yet there is a strong possibility that open reviews (reviews that are signed) are of better quality which would suggest that the reviews used for this thesis also have a better quality as they were all signed reviews.

Summarized, the different research on bias show that removing the names from the reviews will reduce bias. Some bias will still exist as the evaluator might still recognize the reviewer through cultural differences in writing and through recognition of writing style. The fact that the reviewers knew the book authors will have improved the average quality of the reviews, which does not affect this thesis as long as there are still differences in quality to be measured. The authors having knowledge on the identity of the reviewers will also have improved the average quality of the reviews, but as with reviewers knowing the authors this does not affect this thesis as long as differences in quality between the reviews remain.

Expert knowledge

Searching for literature on peer review, certain authors are found frequent. These can therefore be seen as experts on the subject. In the case of Medical Journals one such expert is John Overbeke, a surgeon and Chief Editor of the Dutch Medical Journal (Nederlands Tijdschrift voor Geneeskunde). This expert was contacted through e-mail hoping to become a source of information on possible methods on analysis of peer reviews. However, Mr Overbeke did not know of any such existing methods and advised the following: “contact editors of GIS magazines and ask them about guidelines for book review. Take from these guidelines the most important criteria and use these for your analysis. Do not only describe the contents.”.

Because an expert on the subject does not know about review evaluation methods, it can be assumed not many, if any, methods actually exist. However, a few methods have been found and therefore the need to contact editors of magazines is as yet postponed. In creating new methods however, the expert advice is used. The criteria found to be most important of existing methods are taken and combined into a new method.

Review evaluation instruments

The following list evaluates methods for review evaluation found in scientific articles. After this a list of methods found in semi-scientific literature is evaluated.

- A. Feurer *et al.* (1994) describes a research that tries to “measure the reliability and preliminary validity of a grading instrument for editors to evaluate the quality of peer reviews”. The instrument uses basic attributes (such as number of reviewed sections, timeliness and referencing) of review content and format. Grades have been assigned to each attribute, and the total of these grades result in a final score. This score then reflects the quality of the review (“0” being the lowest and “14” being the highest possible score). The attributes and their respective possible scores can be seen in the following figure.

Grading Instrument and Scoring System

Attribute (Contribution, %)	Points	Scoring Guide Examples
Timeliness (21)*	3	<3 wk
	2	3-4 wk
	1	5-8 wk
	0	≥9 wk
Grade sheet (7)*	1	Completed
	0	Not completed
Etiquette (7)	1	No writing on manuscript; reviewer not identified on narrative; comments not acrimonious, transmissible to author
	0	Good reviewing etiquette not observed
Section-by-section review (21)	3	Each section addressed in narrative review
	2	Some but not all sections addressed in narrative review
	1	Overview provided, but no section-by-section narrative
	0	No narrative review or letter only
Offering supporting references (14)	2	References specifically cited or copies attached
	1	References mentioned
	0	No references mentioned
Summary and/or recommendation (14)	2	Valuable, logical, accurate
	1	Indecisive, not logical, inaccurate
	0	No summary or recommendation provided
New insights/perspectives (14)	2	More than one new insight/perspective provided
	1	One valuable insight/perspective provided
	0	No new insights/perspectives provided
Total possible	14	

*Pre-coded attributes.

Figure 2.1: Peer review scoring system (Feurer *et al.*, 1994)

The possible scores given to the different attributes are related to the importance of each attribute. The importance of the attributes however is arbitrary and based on the designed function of the grading instrument: to help editors evaluate manuscripts reviews of the *Journal of Vascular and Interventional Radiology (JVIR)* along two basic dimensions: content and format. Whichever attribute is seen as more important for this journal receives higher possible scores.

- B. Rooyen *et al.* (1999) describes the so-called Review Quality Instrument (RQI) for assessing peer review of manuscripts. It is of identical set-up as the method by Feurer *et al.* (1994). There are 8 attributes of reviews that can be given a grade. In this method the grading itself is more detailed, having an ordinal scale from 1 (bad) to 5 (good). Between attributes there is less

specificity, as all attributes receive the same weight. The method can be seen in the following figure 2.

Review Quality Instrument				
Paper number:.....	Ref:.....	Referee No:.....	Editor:.....	
1. Did the reviewer discuss the importance of the research question?	1	2	3	4
Not at all				5
				Discussed extensively
2. Did the reviewer discuss the originality of the paper?	1	2	3	4
Not at all				5
				Discussed extensively with references
3. Did the reviewer clearly identify the strengths and weaknesses of the method (study design, data collection and data analysis)?	1	2	3	4
Not at all				5
				Comprehensive
4. Did the reviewer make specific useful comments on the writing, organisation, tables and figures of the manuscript?	1	2	3	4
Not at all				5
				Extensive
5. Were the reviewer's comments constructive?	1	2	3	4
Not at all				5
				Very constructive
6. Did the reviewer supply appropriate evidence using examples from the paper to substantiate their comments?	1	2	3	4
No comments substantiated				5
				All comments substantiated
7. Did the reviewer comment on the author's interpretation of the results?	1	2	3	4
Not at all				5
				Discussed extensively
8. How would you rate the quality of this review overall?	1	2	3	4
Poor				5
				Excellent

Figure 2.2: Review Quality Instrument (RQI) for assessing peer review of manuscripts (Rooyen *et al.*, 1999)

Downside of the RQI is the focus it has on research. The attributes chosen for this instrument apply well on research, but are less interesting for this thesis. This is due to the function the method was designed for, evaluating research article reviews. In the book reviews used for this thesis, no actual research is used, therefore result interpretation and research question importance are not relevant, though similar attributes of chapter reviews can be thought of.

C. The peer reviews that will be analyzed in this thesis have one obvious special attribute: they are not paper reviews, they are book chapter reviews. This might require additional or special attention in the analyses or a different method of analyses altogether. The reviews analyzed in this thesis are no standard book reviews. They do not encompass a story written by an author who read a book, they are reviews of single chapters and they have been done using a standard format for reviewing papers of a scientific journal. Besides these differences they are still book reviews and could be analyzed as such. On the subject of scientific book reviews one topic has been of much debate: the level of scholarliness (scientific level) of these reviews.

A recent paper on this matter was written by Nicolaisen (2002). He proposes a simple method for measuring the scholarliness of book reviews. He first analyses previous methods and as such comes up with several indicators of scholarliness:

- Trustworthiness
 - Publication counts
 - Citation counts
 - Textual Indicators

- Colonic titles
 - References
- Published/Non-published
- Reliability of the author
 - Expertise in the field
- Quality of the review
 - Description of the book
 - Description of the subject
 - Answering of the question: “Is the book an addition to our knowledge, and if so, what exactly has been added?”
 - Position the book in relation to former literature on the subject

Nicolaisen then argues that reviews are hardly ever cited, and colonic titles are irrelevant as book reviewers usually copy the titles of the reviewed books. He does not provide any counter-arguments for author reliability or review quality, yet the reasons for this can be imagined. Nicolaisen seeks an easy methodology for testing the scholarliness of book reviews, in order to be able to assess the quality of such a review in little time. Author reliability requires insight in the field of expertise while review quality requires the reading of each review before being able to evaluate its scholarliness. Therefore Nicolaisen concludes that: “book reviews that are published, accessible and that bear references to additional literature reflect scholarly qualities.”.

Of the three attributes concluded by Nicolaisen to evaluate scholarliness of book reviews, only the bearing of references is relevant for this thesis. None of the reviews will be published, as they are made by students as assignment for a course. So if applying this method when selecting reviews to create a review summary, only reviews that contain references to other literary work should be used as only those reviews are considered scholarly.

Unofficial assessment instruments

As before mentioned, some of the found evaluation methods are semi-scientific. They have a scientific background, yet have not been published in a journal article. Because instruments for peer review are not commonly found in scientific journals, 2 methods found on the internet will now be described. Both are however created on university level as tools for certain courses.

- D. The first described method taken from course materials of Virginia Tech (2007) is a guideline for students showing attributes of reviews and the amount of points that can be scored on these attributes. The list is made for grading reviews of proposals and not of books or articles, and attributes have been chosen accordingly. A maximum of 100 points can be scored and attributes can receive a maximum of 10 to 25 points, depending on the attribute. Detailed point attribution is however not stated. Merely a maximum score is given for a group of attributes belonging to larger categories such as content or structure. The specific amount of points given is up to the analyst of the reviews. For example:

Content (15 points)

- good choice of issue as subject
- reflects knowledge of subject
- length is in assigned range (deduction if word count not specified)
- not repetitive

Content can receive a maximum score of 15 points and the analyst should base his points given on choice of issue, knowledge of subject, length and repetition. The exact point distribution for each of these four attributes is not specified. For total grading sheet see appendix B (Virginia Tech, 2007).

- E. The second semi-scientific method is the so-called “Critical Paper Review Grading Sheet” (University of Massachusetts Lowell, Dept of Computer Science, 2007) It is relatively simple and even more subjected to the impressions of the analyst. The maximum number of points earned is once again 100. Yet the division of these 100 points is limited to 6 attributes (summary of paper, critique, filling out of form, formatting and appearance, grammar and spelling) varying from 5 to 50 points. Point division is more specific for these attributes, yet only in a subjective way, such as 20 points for a poor summary, 40 for average and so on. The exact point distribution can be seen in the following figure 3.

Content (max 85 pts):	
Summary of paper:	
Outstanding detail (50 pts)	_____
Good detail (45 pts)	_____
Average detail (40 pts)	_____
Poor summary (20 pts)	_____
No summary or no submission (0 pts)	_____
Critique	
Outstanding—strengths, weaknesses, reader’s position (35 pts)	_____
Good (30 pts)	_____
Average (25 pts)	_____
Poor (20 pts)	_____
No critique or no submission (0 pts)	_____
Grammar and writing (max 15 pts):	
First two lines of this form filled out, and form <i>stapled</i> to front of review (5 pts)	_____
Formatting and appearance (5 pts)	_____
Grammar and spelling (5 pts)	_____
Total (max 100):	_____

Figure 2.3: Critical Paper Review Grading Sheet (University of Massachusetts Lowell, Dept of Computer Science, 2007)

Review methodology options

It appears that all instruments that have been developed for analysing the quality of peer reviews use a simple grading sheet system. Points are allocated to certain attributes of peer reviews and the eventual analyses provides a grade that reflects the overall quality of the review. On an even simpler scale, one study uses only three attributes for assessing review quality, of which only the use of references to other literature could be used for this thesis.

As most of the methods use a grading sheet, one of the chosen methods will be such a grading sheet. The grading sheet chosen must have attributes that apply to the review format of the book reviews. Furthermore the point allocation must be described as detailed as possible to minimize possible evaluator bias. To test differences between the methods, the other methods are chosen/created in order to have a gradient in time necessary to use the method. This way the eventual testing will show whether fast methods can evaluate review quality as well as more time consuming methods, or whether invested time pays off.

Method 1: Separating reviews that use references from those that do not and then summarizing the suggestions given in those reviews. This method is based on Nicolaisen (2002). This is the least time consuming as those reviews that do not contain references are not even read, they are dismissed immediately.

Method 2: Using a known grading instrument is the second method. The system of Feurer *et al.* (1994) is chosen because this system provides a very specific grading system with uneven allocation of points between attributes. It also has the most attributes that apply to the chosen review format except for timeliness.

Method 3: Incorporating the different grade sheets, creating one grade sheet with attributes that best apply to the review format of the book reviews. Creating this method will be more time consuming than using a known grade sheet, but because the method is designed for the used reviews it is likely the method will better evaluate the quality of the reviews.

Method 4: The last option, not described or used in researches would be to individually assess all comments given in each review. Each comment will have to be assessed on correctness and usefulness and the eventual summary will be a list of the best comments from all reviews. This method will likely be the most consuming but should incorporate all the best comments.

Using these four methods for the reviews of a single chapter should reveal the differences in time spend on the analyses and the differences in quality between these analyses. The eventual choice of best method will be the method that indeed shows the quality of a review using the least amount of time. This can be useful for editors all over the world, as such a method could greatly reduce the time spend on the reviewing process. Δ Time versus Δ quality. It will be interesting to see whether the work of designing a new grade sheet, based on existing grade sheets, will improve the results so much that this becomes worth the effort.

The creation process of methods 3 and 4 will now be explained.

Creating Method 3: Grade sheet for analysing book reviews

The basis for this grade sheet will be the grade sheet made by Feurer *et al.* (1994).

This grade sheet gives specific weights to attributes, each weight reflecting relative importance of the attribute, and has attributes that relate to the chosen review format of this thesis. Furthermore the attributes have very specific guidelines for grading and more importantly these guidelines are non-arbitrary. For instance the attribute of new insights/perspectives has three grading options and each one is very specifically described and not arbitrary like as with some other methods.

One attribute however does not apply and will therefore be removed: Timeliness. All reviews have been collected for an assignment and therefore all reviews were on the same time. Leaving this attribute would provide the same amount of points to each review making the rest of the attributes relatively less impacting.

The next found grade sheet, that of Rooyen *et al.* (1999), has some questions that apply specifically to papers and are not relevant to our book reviews. Some questions however do apply and are not present in the grading sheet of Feurer *et al.* (1994). These questions are:

- Did the user make specific useful comments on the writing, organisation, tables and figures of the manuscript?
- Did the reviewer supply appropriate evidence using examples from the paper to substantiate their comments?

The first attribute will be limited to comments on tables and figures as comments on other parts are already dealt with by other attributes. This attribute will receive three possible weights: 0 (no comments given on figures and tables), 1 (comments given on some figures and tables) and 2 (comments given on all figures and tables).

The second attribute will receive 3 possible weights: 0 (no examples from the book chapter given), 1 (some examples given from book chapter) or 2 (examples from book chapter given to (almost) all comments).

There is also a question on the constructiveness of the comments, however rating constructiveness (the way in which a review helps to improve a reviewed book/article) is very arbitrary and will therefore be left out.

The un-published methods of Virginia Tech (2007) and University of Massachusetts Lowell, Dept of Computer Science (2007) are a lot less specific. The one of University of Massachusetts only assigns such arbitrary and vague attributes that this grade sheet will be unused in the process of making a new grade sheet. It is extremely difficult to scientifically rate a summary for instance, especially if one has to choose between poor, average, good or outstanding, the possibility of several people choosing the exact same is almost nil.

The grade sheet of Virginia Tech is also very unspecific in grading attributes and many attributes can only be graded arbitrarily. One attribute however can be given specific weights and are therefore useful.

- Amount of typos. Other typing and spelling attributes are suggested as well, but only with amount of typos can one give weights non-arbitrarily.

This attribute will be given only two possible scores: 0 (more than 10 typos per page) or 1 (less than 10 typos per page). The maximum weight is not high for this attribute

as it only generally reflects the amount of effort spend on the review but could also be due to lingual problems of the reviewer.

The final grading sheet for method 3 is therefore the following:

Attributes	Points	Scoring guide examples
Grade sheet	1	Completed
	0	Not completed
Etiquette	1	No writing on manuscript; Reviewer not identified on narrative; comments not acrimonious, transmissible to authors
	0	Good reviewing etiquette not observed
Section by section review	3	Each section addressed in narrative review
	2	Some but not all sections addressed in narrative review
	1	Overview provided, but no section by section narrative
	0	No narrative review or letter only
Offering supporting references	2	References specifically cited or copies attached
	1	References mentioned
	0	No references mentioned
Summary and/or recommendation	2	Valuable, logical, accurate
	1	Indecisive, not logical, inaccurate
	0	No summary or recommendation provided
New insights/perspectives	2	More then one new insight/perspective provided
	1	One valuable insight/perspective provided
	0	No new insights/perspectives provided
Comments in tables/figures	2	Comments given on all figures and tables
	1	Comments given on some but not all figures and tables
	0	No comments given on figures and tables
Examples from book chapter	2	Examples from book chapter given with all comments
	1	Examples from book chapter given with some comments
	0	No examples from book chapter given with comments
Typos	1	Less then 10 typos per page of narrative review
	0	More then 10 typos per page of narrative review
Total score possible	16	

Figure 2.4: Grading sheet for analysing book chapter reviews based on "Peer review scoring system" (Feurer *et al.*, 1994)

Creating Method 4: Methodology description for analysing book chapter reviews per comment

The different methods described before that analyse reviews all have one thing in common: they focus on speed. They try to separate quality reviews from inadequate reviews in as little time as possible. It could be argued that each review might provide valuable clues for the authors and that even the best reviews might still have some comments that are false or at least less well thought through. Therefore defining a method that will not separate quality reviews from bad reviews, but separates quality comments from bad comments seems logical.

To make such a method work, each comment must be separated. Then copy each comment into a pre-made form in Microsoft Excel where certain weights, attributes or characteristics can be given to each comment. The total of these attributes should result in the usefulness of the comment. Useful comments can then be selected and used to create the eventual summary that is provided to the book authors as a recommendation for improvement.

Though multiple attributes exist that make a comment useful, in order to create a fast method attributes must be defined that certify the usefulness of a comment in as few attributes that cost the least amount of time and effort as possible. One un-arbitrarily identifiable attribute is objectiveness; though a subjective comment is not by definition bad, a factual book will improve more by adding new or falsifying old facts instead of changing opinions of people.

A second attribute is the number of times the comment is given. If several reviewers give the same comment, this points towards an obvious flaw in the chapter and should therefore be seen as a valid comment. Even if this comment is subjective, if several people agree with the comment, it should at least be taken into consideration.

In the case of objective comments, the availability of a reference is another attribute that can verify a given comment. Though each reference must be examined on source and age. If the reference is already present the chapter authors will have a reference for their work and therefore be saved of searching for one. It verifies the comment as a fact and thus the comment can be used.

These three attributes select comments on factuality through referencing, objectiveness and numbers through count. However it does not de-select comments that have all these attributes yet do not seem to have any connection to the actual book chapter. Comments for instance on adding facts about technology to a chapter about data might be objective and referenced, yet out of place. There is no easily identifiable attribute for this problem and such comments will therefore have to be identified by common sense and subject knowledge, this attribute will be called "relevance".

The following flow chart shows the decision procedure used for this method. First objectiveness is assessed for a comment. Objective comments are then assessed on referencing, followed by count and relevance, while subjective comments are directly assessed on count and then on relevance.

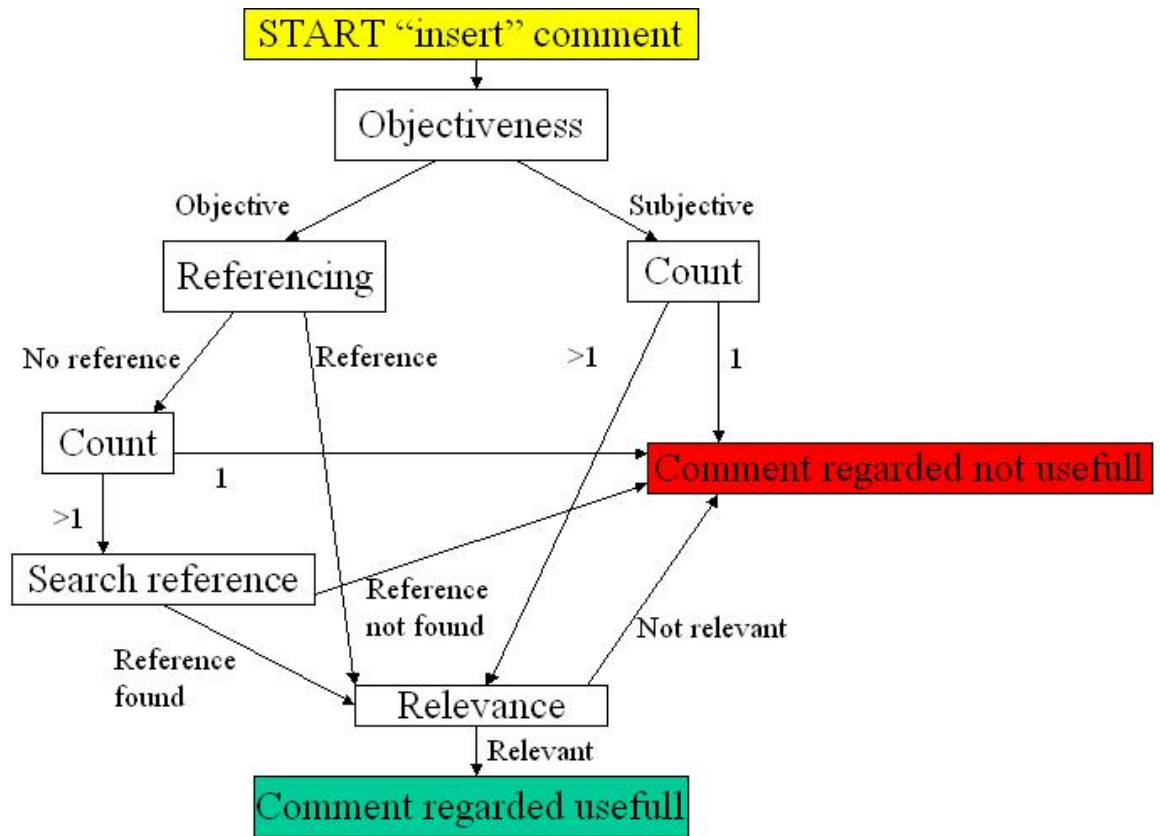


Figure 2.5: Decision flowchart for method 4

Methodology for stepwise analyses of the results

Analyses of the results is done in two mayor steps. First methods 1, 2 and 3 are compared to a baseline provided by the book authors. Method 4 is then compared to the best method(s) from this step. This second step is necessary due to the different nature of the results gotten from method 4. The total analyses process is shown in the figure below.

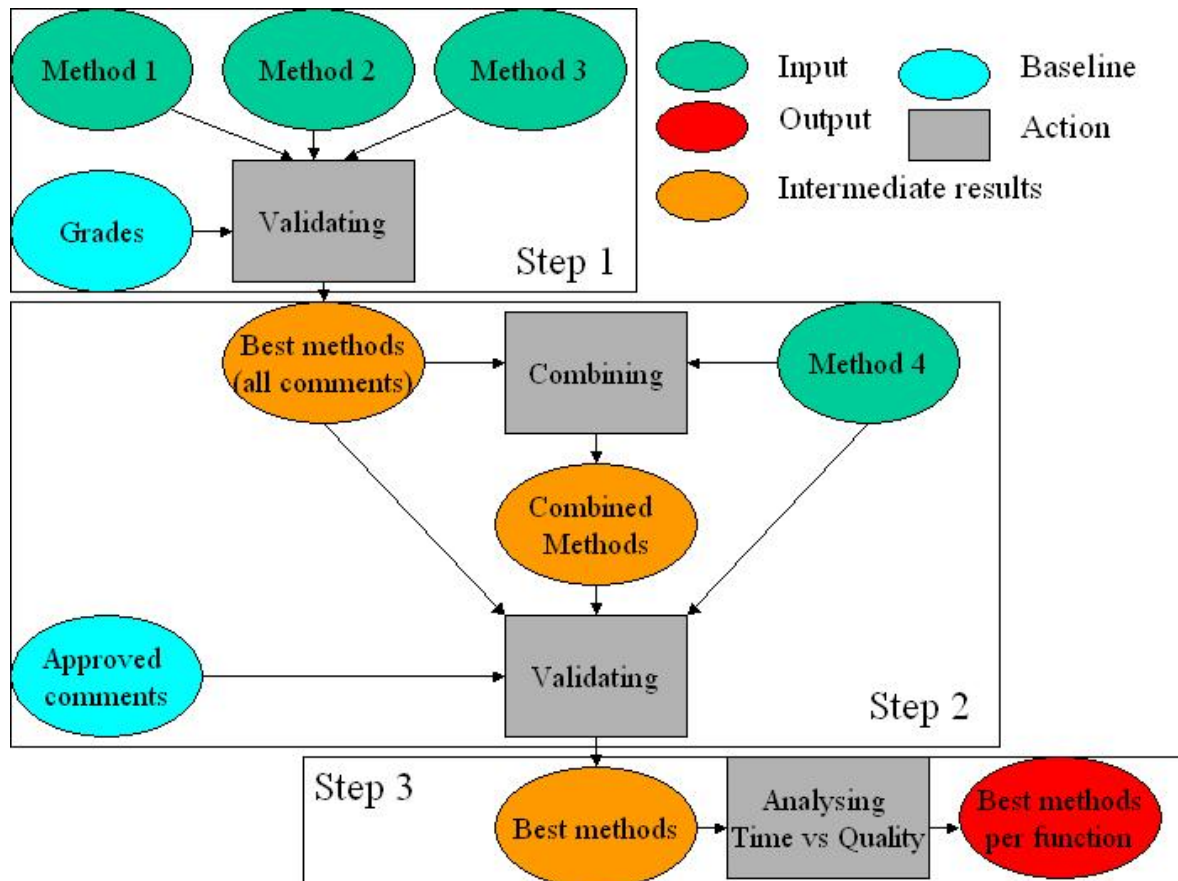


Figure 2.6: Methodology scheme for result analyses

Step 1: Methods 1, 2 and 3 compared to author's grades

For the analyses of methods 1, 2 and 3, grades given by the book authors are used as a comparison baseline. For analyses all three methods must have their results converted to similar formats. As method one yields only acceptance/rejection results, the results of methods 2, 3 and the baseline itself must be converted to the same result format.

This conversion will be done in 2 ways. Firstly by accepting all reviews that have a grade equal to or higher then 5.5. This grade is chosen because 5.5 is the generally accepted grade in Dutch higher education for a certain test object to be sufficient. Second conversion will use a grade of 7.0 (the grade that marks the level of "More than sufficient") as accepted reviews. The reason for this is that it can be argued that strictly sufficient can only be seen as such for testing in a school environment. And for material to be used in scientific research a higher level is to be expected.

After conversion the percentage of similarity is measured of each method compared to the grades baseline. The higher the percentage is, the higher the similarity of the method compared to the baseline and therefore the better the results.

Methods 2 and 3 can have more extensive analyses through statistic analyses. Correlation can show whether the methods show a similar trend as the baseline and linear regression can show exactly which trend this is.

The percentage of similarity mentioned earlier is more important as method 1 can not be converted to grades and therefore only the percentage of similarity analyses can be used to compare all three methods to the baseline in a similar way.

Step 2: Separate comments of method 4 and of the best methods of step 1 compared to author's accepted separate comments

For the second step again conversion is needed of the separate methods in order to be able to make a sound comparison. Method 4 already separated accepted comments from rejected comments. The methods that step 1 show are acceptable methods (showing better than random results) have to be converted to the same format. There will be two conversions. One is to take all comments from the accepted reviews. The second is to only take the comments that would be accepted through method 4 of the accepted reviews (accepted using the 5.5 border grade). This will yield less comments than method 4 itself, because method 4 bases acceptance of comments on the number of times the comment is mentioned. When through another method less comments are considered, chances of separate comments being accepted decline.

As baseline the authors have provided a list of all the comments and checked which comments they would use for improving their book. In order to do this the authors have taken the number of times a comment was given into account as well as whether a reference was provided with the comment.

Similar to step 1, the similarity between the baseline and the methods is then measured as a percentage. The higher the percentage, the higher the similarity in results between the methods and the baseline and therefore the better the results.

Step 3: Final analyses of the methods comparing quality with time consumption

The final analyses is a comparison of all the methods accepted after step 2 of the analyses process. For each method total time is counted and compared to quality (percentage of similarity with the baseline).

In this comparison there is no "good" or "bad" method. The key point is to show the differences in time consumption and quality yield of each method. For any given objective a certain quality is expected of a review. The relative time consumption can be important when selecting a method for certain objectives. After comparing quality with time consumption conclusions can be made on the usefulness of each method for different objectives.

Chapter 3: Results & Interpretation

The results will be presented in a stepwise analyses. The steps used are the same that can be found in figure 2.6 at the end of the methodology chapter. After the final step one or more methods will have been selected. These methods will then be used to create the second part of the results: presenting the main suggestions for improving the book chapters. Detailed results data per review can be found in Appendix C.

Step 1

Step 1 of the result analyses can be divided into several smaller steps. Each of these smaller steps will be indicated by a letter in alphabetical order. Smaller step can be divided into even smaller parts. This division is necessary to ensure the understanding of the stepwise progress of the analyses.

Step 1A: Methods 1 through 3 compared to baseline using acceptance/rejection results

The first three methods all produce results that can be translated in either acceptance or rejection of each review. For method 1 this acceptance or rejection is definite. For methods 2 and 3 the acceptance or rejection is shown through a grade between 0 and 10. The same type of grading is used for the baseline. To be able to compare the results of methods 2 and 3 (and the baseline) with the results of method 1, each grade must be translated to either concrete acceptance or rejection.

The graded methods will be translated twice. Once using 5.5 or higher (the standard Dutch university level of acceptance) and once using 7.0 or higher (the grade that marks the level of “More then sufficient”) as accepted reviews. The reason for this is that it can be argued that strictly sufficient can be only be seen as such for testing in a school environment. And for material to be used in scientific research a higher level is to be expected.

The following table shows the similarity in acceptance (shown in %) between the three methods and the baseline. The higher the similarity, the better the quality of the method.

Acceptance grade	Method 1	Method 2	Method 3
Acceptance if grade => 5.5	61%	86%	71%
Acceptance if grade => 7.0	54%	68%	68%

Table 3.1: Percentage of similarity in acceptance of reviews between method 1 through 3 and baseline using two levels of acceptance.

The table clearly shows that method 1 is of unacceptable quality. The percentages are near 50 % which (because of the fact that only 2 options, acceptance or rejection, are faced in this test) indicates randomness of the results. It will therefore not be taken into account in the final conclusion that will reflect quality versus time spend.

Methods 2 and 3, the methods similar to each other, show results that are better then random for both levels of acceptance. It is striking that for a border grade

of 7.0 both methods yield the exact same results. Yet when applying the lower 5.5 border grade, method 2 yields better results. In fact the results of method 2 are nearing 100 % , and are 15 % better then those of method 3.

Step 1B: Methods 2 and 3 compared to baseline (grades) using actual grades

Part 1: Basic attributes

The basic attributes can show differences in averages and standard deviations. The following table shows these attributes.

Method	Average	Median	Standard Deviation
Method 2	7,3	7,1	1,72633708
Method 3	6,4	6,3	1,746003562
Grades	7,5	7,4	1,357128933

Table 3.2: Basic attributes of methods 2, 3 and the baseline (grades)

The average and median show that method 2, on average is closer to grades then method 3. Both methods 2 and 3 have similar standard deviations, which indicates that both methods have a similar pattern in extremes. The standard deviation of the baseline is lower (1.35 versus 1.73 and 1.75 of methods 2 and 3) then that of methods 2 and 3 which indicates that both methods have stronger extremes then the baseline.

Part 2: Correlation

In order to see whether a trend exists between the methods and the baseline the linear correlation is calculated. Only linear correlation is tested because if the results are perfect the linear trend $y=x$ is found. The following table shows the linear correlation.

Method	Correlation with grades
Method 2	0,648630737
Method 3	0,777193565

Table 3.3: Linear correlation between methods 2 and 3 and the baseline (grades)

Method 3 is shown to have a higher correlation then method 2 and therefore shows a stronger trend then method 2.

Part 3: Regression analyses

Even though correlation has shown that method 3 has a stronger linear trend then method 2, only regression can show what trend this actually is as linear correlation does not guarantee a $y=x$ trend. In the following figure therefore methods 2 and 3 have been plotted against the baseline grades. For both methods a trend line is plotted as well with trend formula and R^2 . For both methods the regression has a significance below 0.05 and therefore both regression analyses are significant.

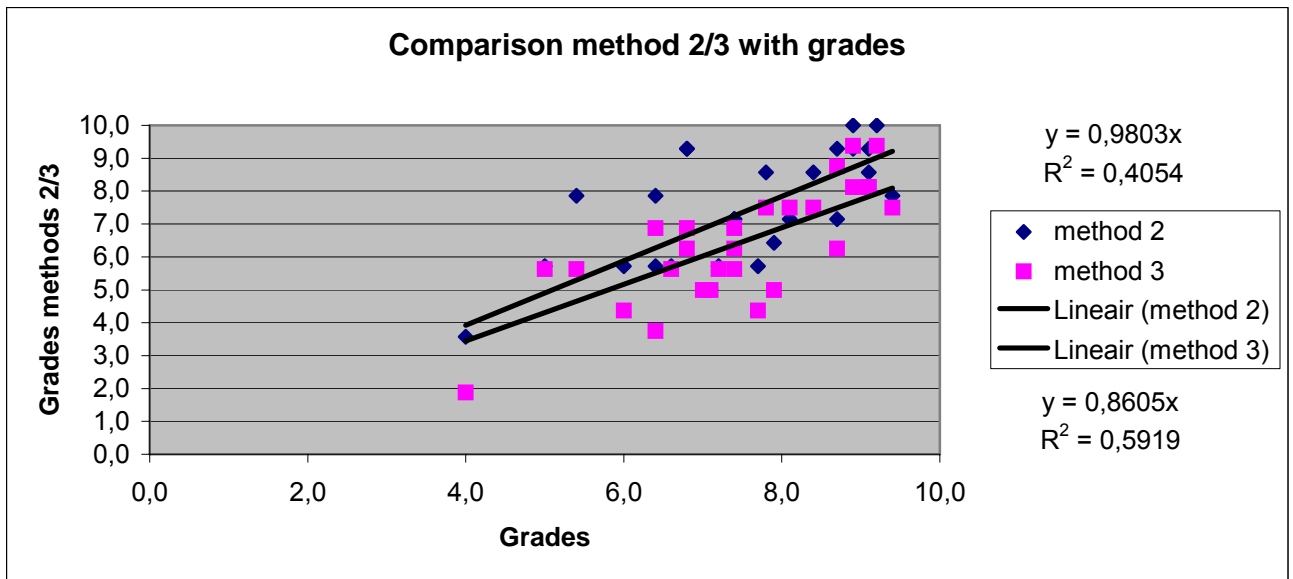


Figure 3.1: Methods 2 and 3 plotted against baseline (grades) with added linear trend lines.

The figure shows that for method 2 the formula of the trend line is $y=0.9803x$ which nears the perfect trend line $y=x$. For method 3 the plotted trend line is $y=0.8605x$. This indicates that, on average, each grade of method 3 is 1.4 lower than the baseline as expected when analysing average and median.

The R^2 of method 2 = 0.4054 and for method 3 the measured $R^2 = 0.5919$. This means that the unexplained fraction of the trend line is larger for method 2 than it is for method 3, as expected by the higher correlation of method 3 with the baseline.

Step1 C: Calibration of method 3

The explained fraction of method 3 is higher than that of method 2, yet the trend line of method 3 has a constant deviation from the expected $y=x$. Therefore method 3 requires calibration.

Because the expected trend line is $y=x$, calibration of method 3 takes only 1 step. All grades given by method 3 have to be multiplied by $1/\text{trend line fraction}$ (0.8605), in this case approximately 1.16. This calibration does imply a possible deviation from the fixed set of rules, which is grades surpassing the border grade of 10.

Part 1: Methods 1, 2 and calibrated method 3 compared to baseline using acceptance/rejection results

Method	Method 1	Method 2	Calibrated method 3
Acceptance if grade => 5.5	61%	86%	82%
Acceptance if grade => 7.0	54%	68%	68%

Table 3.4: Percentage of similarity in acceptance of reviews between method 1, method 2, calibrated method 3 and baseline using two levels of acceptance.

The above shown table X shows the percentage of similarity of acceptance/rejection of the methods compared to the grades baseline. Methods 1 and 2 are unchanged and so method 1 still nears random and is therefore discarded

as possible method. Method 3 is unchanged for the border grade of 7.0 and remains the same as method 2. For the border grade of 5.5, the percentage of method 3 has increased to 82%, nearing the 86% of method 2. Method 2 and 3 therefore now yield similar results using this analyses.

Part 2: Basic attributes

Once the calibration is applied, attributes of method 3 change. They can be found in the following table along with the unchanged attributes of method 2 and grades for comparison.

Method	Average	Median	Standard Deviation
Method 2	7,3	7,1	1,72633708
Calibrated method 3	7,4	7,3	2,028978785
Grades	7,5	7,4	1,357128933

Table 3.5: Basic attributes of methods 2, calibrated method 3 and the baseline (grades)

Average and median of calibrated method 3 now, as expected, near the average and median of grades. The standard deviation of method 3 has also increased. Because of the multiplication with a factor of 1.16, the extremes of method 3 are also increased.

Part 3: Correlation and regression of calibrated method 3

Because the calibration consists only of a multiplication of all the grades, correlation remains the same as found in table x of results part 2B (0.77). The effect of the calibration can be seen in the regression (see figure X below). The trend line now follows the expected $y=x$. The explained fraction remains the same.

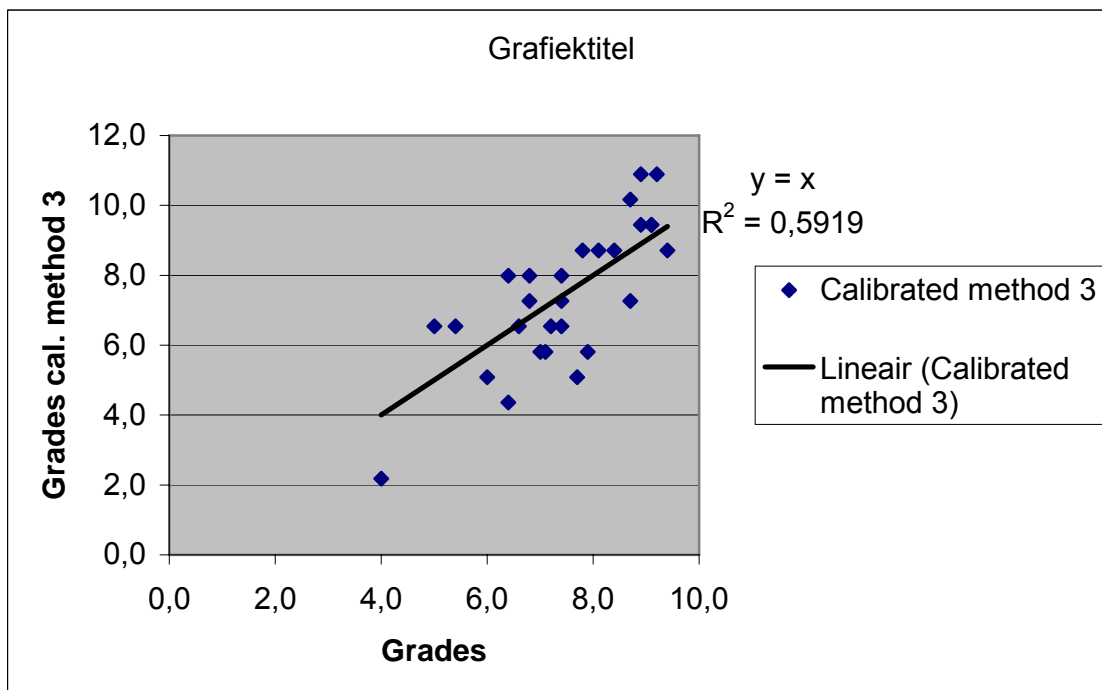


Figure 3.2: Calibrated method 3 plotted against baseline (grades) with added linear trend line

Step 2

The second step of the research no longer focuses on separate reviews. Instead the focus is on each separate comment found in the different reviews. The newly made fourth method uses a system that either accepts or declines the use of an individual comment for improving a book chapter.

A list with all the comments was provided to the two book authors. They were asked to indicate which comment they would use for the improvement of their book. Each author was given only the comments, the amount of times the comment was given by reviewers and any references given with the comments. The comments they selected form the baseline for this phase of the research. The comments only one of them selected were not used, only those selected by both authors. This is done to select only the best comments.

Methods 2 and the calibrated method 3 were chosen from the first step as best possible methods for analysing book chapter reviews and therefore these two methods will be used in the second step of the research. Each method will be combined with method 4, resulting in selected comments from the reviews chosen by methods 2 and calibrated method 3.

In addition, method 4 alone will be compared to the combination methods, as method 4 on itself already picks separate comments. Method 2 and calibrated method 3 can also be used without combining them with method 4, effectively choosing all comments from the reviews chosen by these methods. However as the second step analyses the separate comments of the methods, the full reviews of method 2 and calibrated method 3 will not be taken into account in these analyses (they would simply yield a 100% score). These options will be taken into account in the final analyses and conclusion. In this last step it is weighed whether selecting separate comments through method 4 or combinations of other methods with methods 4 is faster then selecting full reviews as done in the first step of the research.

The following three tables will show the amounts of comments that were given on each chapter, as well as the amount of comments chosen from these comments by the separate authors and both authors. In green will be shown the amount of comments from each of these categories that were accepted by method 4 and in red the amount of comments that were rejected by method 4. The first table shows all comments, while the second and third table show only the comments from the reviews chosen by method 2 and calibrated method 3 respectively.

As the tables can seem complex, a small example will now be highlighted for clarification.

Chapter	All	Bregt	Crompvoets	B&C
Chapter 1	24	13	17	10
	3	3	3	3
	21	10	14	7

In this part of the table, only the results for chapter 1 are shown. The first line, in black indicates all the comments (All) from reviews of chapter 1. Followed by those selected by Mr. Bregt (Bregt) and those selected by Mr. Cromvoets (Crompvoets). The last column indicates the number of comments that were selected by both Mr. Bregt and Mr. Crompvoets. The second line, in green, shows comments in the same order. These numbers indicate the number of comments from the total that method 4

selected to be used for improving chapter 1. The third line, in red, shows the exact opposite as it indicates those that were not selected. By summing up the red and green numbers, the total number is once again reached, the black number.

Chapter	All	Bregt	Crompvoets	B&C
Chapter 1	24	13	17	10
	3	3	3	3
	21	10	14	7
Chapter 2	44	16	31	11
	13	5	13	5
	31	11	18	6
Chapter 3	37	18	27	16
	3	2	3	2
	34	16	24	14
Chapter 4	51	26	31	24
	13	12	12	11
	38	14	18	13
Chapter 5	41	22	31	17
	9	5	8	5
	32	17	23	12
Chapter 6	35	23	18	14
	8	7	7	6
	27	16	11	8

Table 3.6: The total amount of comments per chapter, as well as those chosen by the separate authors and by both authors.

Chapter	All	Bregt	Crompvoets	B&C
Chapter 1	22	11	17	10
	3	3	3	3
	19	8	14	7
Chapter 2	44	16	31	11
	13	5	13	5
	31	11	18	6
Chapter 3	37	18	27	16
	3	2	3	2
	34	16	24	14
Chapter 4	37	20	23	19
	12	11	11	10
	25	9	12	9
Chapter 5	41	22	31	17
	9	5	8	5
	32	17	23	12
Chapter 6	35	23	18	14
	8	7	7	6
	27	16	11	8

Table 3.7: The comments per chapter of the reviews chosen by method 2, as well as those chosen by the separate authors and by both authors.

Chapter	All	Bregt	Crompvoets	B&C
Chapter 1	24	13	17	10
	3	3	3	3
	21	10	14	7
Chapter 2	44	16	31	11
	12	5	12	5
	32	11	19	6
Chapter 3	37	18	27	16
	3	2	3	2
	34	16	24	14
Chapter 4	51	26	31	24
	13	12	12	11
	38	14	18	13
Chapter 5	40	22	31	17
	7	5	6	5
	33	17	25	12
Chapter 6	35	23	18	14
	8	7	7	6
	27	16	11	8

Table 3.8: The comments per chapter of the reviews chosen by calibrated method 3, as well as those chosen by the separate authors and by both authors

The main analyses of these tables results in the following table. This table shows the percentages of comments that were accepted by the different methods or method combinations and approved by both authors compared to the total amount of approved comments by the authors. The higher these percentages, the better the method, as 100% indicates that the method accepted all comments that were approved by the authors.

Chapter	m4	m2+m4	cm3+m4
Chapter 1	30%	30%	30%
Chapter 2	45%	45%	45%
Chapter 3	13%	13%	13%
Chapter 4	46%	42%	46%
Chapter 5	29%	29%	29%
Chapter 6	43%	43%	43%
average	34%	34%	34%

Table 3.9: Percentages of comments accepted by methods and approved by both authors compared to total amount of approved comments by both authors

On average only 34% of the comments approved by both authors were selected by either method 4 or the combination methods. This means that only one in three comments that should have been indicated was actually selected. This could indicate that method 4 is either too demanding in its selection criteria or that the wrong selection criteria were used.

Second analyses focuses on the comments accepted for use by method 4 or the combination methods. The following table shows the percentages of comments approved by both authors compared to the total amount of comments accepted by the methods. This percentage indicates whether the selecting of comments through

method 4 resulted in a selection (complete or incomplete) of good results or in a random selection of comments.

Chapter	m4	m2+m4	cm3+m4
Chapter 1	100%	100%	100%
Chapter 2	38%	38%	42%
Chapter 3	67%	67%	67%
Chapter 4	85%	83%	85%
Chapter 5	56%	56%	71%
Chapter 6	75%	75%	75%
average	70%	70%	73%

Table 3.10: Percentages of comments approved by both authors compared to the total amount of comments accepted by the methods

The averages indicate that 70% of the comments that were accepted by method 4, were also accepted by the book authors. In combination with method 4 this average increases to 73%. Those comments that are selected by method 4 are therefore mostly usable comments indeed. This indicates that the selection criteria of method 4 indeed do select correct criteria.

Step 3 Analysing methods: quality versus time consumed

In the first phase, method 1 was already discarded due to random results. The results of the second phase showed that on average for method 4, and the combination of method 4 with the other methods, each time only 34% of the comments that were accepted by the authors were also accepted by method 4 or the combination methods. If one wants to fully improve a book chapter, using only one in three usable comments is too little by far. Therefore only taking the full reviews into account will do and for the third step of the analyses method 4 and the combination methods will not be used. This limits the method analyses of time versus quality to methods 2 and calibrated method 3.

Method	Chapter 1 (4)	Chapter 2 (6)	Chapter 3 (4)	Chapter 4 (6)	Chapter 5 (5)	Chapter 6 (4)	Total	Average
Method 2	32.34	42.15	34.27	45.36	60.13	30.15	245.20	8.4
Method 3	40.40	56.14	47.00	56.16	71.37	38.00	309.37	10.6

Table 3.11: The total time needed to analyse the chapter reviews with method 2 and 3 in minutes (between brackets the number of reviews per chapter)

The above table shows the time used to analyse the reviews of the book chapters using methods 2 and 3. On average method 2 takes 8.4 minutes per review. Method 3 takes on average 26 % more time, taking about 10.6 minutes per review.

Regression, done in step 1 of the analyses, showed that method 2 resulted in a good regression ($y=0.98x$) but with a low explained fraction of 0.41. Method 3 had a higher explained fraction (0.59) but the regression had a systematic error resulting in a regression with a linear formula of $y=0.86x$. After calibrating method 3, regression was a perfect $y=x$ with the same explained fraction.

Un-calibrated method 2 had better results and was faster. However method 3 had better results after the method was calibrated, as the explained fraction was 43% higher than that of method 2. Time consumed was 26% higher (excluding calibration time), therefore quality increase is relatively higher than time increase (43% vs 26%).

Method 3 did need calibration, and calibration needs a high number of reviews. Extra time needed for the calibration becomes relatively less the more reviews are analysed with the method. So time consumed becomes relatively less the more reviews are eventually analysed. Method 3 therefore becomes more interesting with higher numbers of reviews.

Presenting a summary of the main suggestions for improving the book chapters

In order to present the main suggestions, one needs to present a selection of the different comments given. The best methods only accepted reviews, not comments. Though it is possible to only select the comments from those reviews, any good comments from the not selected reviews are discarded.

Method 4 only provided on average 34% of the comments that would be accepted by the authors. However, 70% of the comments that were approved by method 4, were also approved by both authors. So for the selected comments there was a high agreement between the authors. This indicates that the comments chosen by method 4 are indeed high quality comments and therefore they are chosen to be the selection used as main suggestions for improving the book chapter. This does imply that other comments are valid as well but not given here as main suggestion, as indicated before, the comments of method 4 only encompass 34% of all the valid comments, yet they can be seen as the core suggestions. The comments selected by method 4 per chapter are the following:

Chapter 1

- The first paragraph is already an example, so to have a second paragraph called examples is unnecessary, choose 1
- Michelin Route example should be made anonymous as the guidelines reject commercial advertising
- The driving forces are mentioned in a different order then in the summary, this should be corrected

Chapter 2

- Give concrete definition of the most important concepts of the chapter
- Start with the definition of Spatial Data
- Add "users" as a paragraph to the chapter
- Introduction does not introduce all the topics
- Old map is not a relevant example for the chapter
- Explain the links and blocks of figure 2.2
- Explain the other core datasets (other then geodetic) in more detail
- Give a wider definition of interoperability
- Treat all concepts in conclusion
- Leave out the new concept of data maintenance and investment in the conclusion
- Add data properties
- Add information on semantic interoperability
- Alter figure 2.4 by adding back loops and splitting data owner and data user/replace figure 2.4 by clearer figure

Chapter 3

- Clarify/remove figure 3.1
- Clarify and list concept of framework such as leadership, long-term vision, communication channels and awareness
- Add to conclusion information of all subchapters

Chapter 4

- Describe some common standards (such as ISO) and their differences
- Add examples for ad hoc and proprietary standards
- Add a conclusion/summary
- Add figure on decision making process for standards
- Select examples of ISO standards instead of giving a long list
- End of page 5 ICT standards are said to be discussed in the next paragraph, however, they are not
- Clarify figure 4.1, applications seems to have the same value as the standards
- Add more information on GML and XML or link to another chapter
- Explain the concepts of different services standards or link to the explanation in another chapter
- Explain standards for people
- Explain standards for licensing
- Add general definition of standards
- Add figure on CEN, ISO and OGC

Chapter 5

- Add an overview of the whole chapter
- Create a subparagraph for both DBSM and DDBSM
- Figure 5.2: This figure is very vague, and doesn't give a clear overview of the 4 components. Next to this, terms like service chaining, direct data access and middleware aren't explained at all or only later in the chapter. This figure doesn't help out in understanding the generic interoperability model; it only makes the technical aspects of an SDI more difficult and confusing, which is the opposite of the purpose of the model.
- Make a new paragraph "standards needed for the SOA" with a list of explanations of XML, GML, SOAP, etc.
- Add a summary that links the paragraphs, and shows a list of all the abbreviations
- Two figures 5.8 Both do not add value and should be removed
- Make new and clearer division of the chapter with new headings
- Explain the term ASI
- Explain Geoportal in more detail

Chapter 6

- Elaborate on the issues of table 6.2 and add them in the text with explanation
- Explain the funding models in more detail (Add advantages and effects of funding types, possibly in a figure)
- Add a conclusion
- Add examples from real world policies
- Use better definition on policies
- Add views from stakeholders and effects that result from different disciplines and cultural backgrounds
- Add "legal aspects of SDI" under a different subtitle
- Place table 6.4 under questions, not in the main text

Chapter 4: Conclusion

In order to reach the main objective of this thesis, “To analyse the student reviews in order to present an overview which will help improve the quality of the book ‘An introduction to Spatial Data Infrastructures’.”, the analyses of the different methods must be transformed into actual comments. To reach this objective each method will be evaluated separately first. Afterwards the main objective will be evaluated followed by discussion and recommendations.

Method 1

Method 1, the method by Nicolaisen (2002), was immediately discarded after the first analyses. The method had been shrunk, as Nicolaisen uses three attributes that indicate a good review, yet only 1 of these attributes, the use of references, could be used for the review format of this thesis. This attribute proved to be too little to indicate a good review. In fact, the method resulted in completely random results. Using this method is therefore dissuaded. The original method using three attributes could be used, only if the review format allows it.

Method 2

Method 2, the method created by Feurer *et al.* (1994), was used in both the first and the second phase of the thesis. In the first phase, method 2 yielded the best results if no calibration (as done in the case of method 3) is done. On top of that, method 2 was also faster than methods 3 and 4. Only method 1, the method that yielded random results, was faster. However, the explained fraction of the regression of method 2 was only 0.41. The method is solid and can be used directly for analysing book chapter reviews. If time is essential in analysing book chapter reviews or if only few book chapter reviews are to be analysed, choosing an existing method (in this case Feurer *et al.* (1994)) is recommended.

Method 3

Method 3, which is the revised version of Feurer *et al.* (1994) using other methods as well, yielded worse results than method 2 in the first phase. The explained fraction of the regression was relatively high at 0.59 (43% better than the fraction of method 2). The worse results were due to a systematic error. To decrease the error the method was calibrated by structurally adding the missing fraction. After calibration, the explained fraction of the regression remained the same and the systematic error was removed. After calibration the results were therefore better than those of method 2. Time wise method 3 took 26% more time than method 2. Improvement of results is therefore relatively almost double that of the extra time cost. If time is less important, improving and calibrating an existing method can be rewarding. The quality improves more than the time consumed. Creating and calibrating the method does take time and initial investment. This time becomes relatively less when the method is used more. If a method will be used many times, improving an existing method is recommended. The calibration can only be done if a few dozen reviews are used. To calibrate the method, these reviews need to be analysed by the original book authors which takes a lot of time and effort. Only those reviews analysed after calibration will profit from this original investment, so it is recommended to make this investment only if the method will be used often, for instance by magazine editors.

Method 4

Method 4 was created from scratch. It is the method that takes the most time, just over double the time of method 2. Method 4 separate and combined with methods 2 and 3 yielded similar results. On average each method of combination accepted 34% of the comments accepted by the authors. Because the results of method 4 and the combination methods were almost the same, and the combination methods take 50% more time, it is recommended to use method 4 on itself and not in combination. If all useful comments have to be found in the reviews, method 4 is insufficient. Only one third of the comments are found. If editors or authors seek to find only main suggestions for improving the book chapters, method 4 can be used. Of the comments selected by method 4, 70% were selected by the original book authors as well. This method still needs development. The choice-scheme created for this method focused on simplicity and speed. In order to improve the results of the method, either by selecting more comments or by selecting less uninteresting comments, the scheme can be adapted. More research can be done to improve this method and can result in a method that selects core comments in reviews. Improvement of the method is especially necessary if core comments need to be selected from single reviews as the choice-scheme partially chooses comments on the amount of times the comments are given in multiple reviews.

Of the four methods, 3 are recommended for further study and use. And the one method that was not recommended was greatly altered for use on this thesis. Each method has different advantages over the other methods and should therefore be used for different purposes.

The initial purpose and incentive for this thesis is described in the main objective. However the book 'An introduction to Spatial Data Infrastructures' is not the only work that can benefit from this thesis. Other book chapter reviews in the future can also be analysed by the methods described in this thesis. And with further research even article reviews done by editors can be greatly improved and simplified.

Main objective: To analyse the student reviews in order to present an overview which will help improve the quality of the book 'An introduction to Spatial Data Infrastructures'.

The authors of the book can improve their book chapters using two methods. The calibrated method 3 has selected good reviews that can be used for improving the book chapters. If the authors are interested in more detailed comments, method 4 has selected core suggestions for improvement.

However during the creation of this thesis, the authors have already selected both reviews and comments that they regarded as useful for the improvement of their book. These were used as base line to which the different methods were compared. The authors therefore no longer need the methods for the improvement of their current book. The methods can be used for improvement of future books by both the authors and other future authors of books.

Discussion & Recommendations

The methods that have been analysed and described in this thesis can be used for more than the review of book chapters. Full book reviews, and even article reviews could potentially be analysed by these methods as well. Slight adjustments, done in the same fashion as with the creation of method 3, can greatly improve the results for different review formats. In any such further study first existing methods should be searched for these formats.

The research done in this thesis is partly based not on previous scientific research, but on logical deduction. The reason for this is that the art of peer review is still in its infant stage. The creation of this work was therefore a creative process. Until the scientific world has a general agreement on what a good review is, the task of assessing review quality seems daunting. Though some methods for this purpose already existed, they were mostly in the form of simple scoring cards. And indeed a scoring card system can certainly point out bad reviews, as this thesis also verifies.

Yet scoring systems are limited in their uses. A scoring system does not regard reviews as a sum of parts, it regards the reviews as a whole. As a result, a review analysed by a scoring system is either accepted and used or rejected and thrown away. To deviate from this trend another method, based entirely on logic, was created. The method tried to view each single comment and evaluate its usefulness. Though not entirely successful in its goal, method 4 did choose good comments and could be the base for further research.

If any such research is attempted, it is recommended that totally new attributes are used initially. Though the attributes of method 4 resulted in fairly good results, the basis for these attributes logical deduction. The common sense of any other can result in equally good or even better attributes and should not be limited by the basis provided here.

It is not recommended that attempts to create new scoring card systems are started from scratch. As many of these systems already exist, at least an overview of different attribute directions can be inferred from them. The focus for a practical system should be on the maximum amount of points that can be awarded to each attribute. The maximum points per attribute will differ per editor and most of all per purpose of each system.

List of cited literature

Baldwin, J., 1998. Putting Your Trust in Reviews: The Ethics of Book Reviewing, Charleston conference, 1998.

Bernard, L., Fitzke, J., Wagner, R., 2004. Geodaten-infrastruktur Grundlagen und Anwendungen. Herbert Wichmann Verlag, Heidelberg, Germany (In German)

Bregt, A., Meerkerk, J., Peersmann, M., Zeeuw, de, K., 2006. Waarheen met de nationale geo-informatie infrastructuur?, GEO-INFO 2006-7/8, Pages 296-301

Crompvoets, J., 2006. National Spatial Data Clearinghouses, Worldwide Development and Impact. Phd. dissertation Wageningen University, 128 pages

Crompvoets, J. and Bregt, A., 2003. World status of national spatial data clearinghouses, URISA Journal, Special Issue on Pullic Participation GIS, 15(APA I), Pages 43-50.

Crompvoets, J. and Bregt, A., 2006 (unedited version). An introduction to Spatial Data Infrastructures.

Feurer, I., Becker, G., Picus, D., Ramirez, E., Darcy, M., Hicks, M., 1994. Evaluating Peer Reviews; Pilot testing of a grading instrument. JAMA. 1994; 272: pp.98-100

Fisher, M., Friedman, S., Strauss, B., 1994. The Effects of Blinding on Acceptance of Research Papers by Peer Review. JAMA. 1994; 272: pp. 143-146

Groot, R., McLaughlin, J., 2000. Geospatial Data Infrastructure: Concepts, cases and good practice. Oxford University Press Inc., New York, US

Hartley, J., 2005. Book reviewing in the BJET: a survey of BJET's referees' and writers' views, British Journal of Educational Technology, Volume 36, No.5, 2005, Pages 897-905.

Jefferson, T., Wager, E., Davidoff, F., 2002. Measuring the Quality of Editorial Peer Review, JAMA, 2002, Volume 287, No. 21, Pages 2786-2790.

Joyce, J., Rabe-Hesketh, S., Wessely, S., 1998. Reviewing the Reviews, The example of Chronic Fatigue Syndrome, JAMA, 1998, 280, Pages 264-266.

Loenen, van, B., 2006. Developing geographic information infrastructures, The role of information policies. Phd. dissertation Delft University of Technology, DUP Science, 390 pages

Nebert, D. (editor for GSDI technical work group chair), 2004. Developing Spatial Data Infrastructures: The SDI Cookbook, version 2.0.

Nicolaisen, J., 2002. Indicators of scholarship; The scholarliness of published peer reviews: a bibliometric study of book reviews in selected social science fields. Research Evaluation. 2002, Vol. 11, No. 3: pp. 129-140

Pierie, J-P, Walvoor, H., Overbeke, A., 1996. Readers' evaluation of effect of peer review and editing on quality of articles in the *Nederlands Tijdschrift voor Geneeskunde*. The Lancet. 1996; Vol. 348, Issue 9040: pp. 1480-1483

Rooyen, Van, S., Black, N., Godlee, F., 1999. Development of the Review Quality Instrument (RQI) for Assessing Peer Reviews of Manuscripts. *J Clin Epidemiol*. 1999; 52, No. 7: pp. 625-629

University of Massachusetts Lowell, Dept of Computer Science, 2007. Critical Paper Review Grading Sheet. http://www.cs.uml.edu/~gary/classes/91_543/Misc/Review_grading_sheet.pdf as seen on February 14th 2007.

Vanderhaegen, M. and Muro, E., 2004. Contribution of a European spatial data infrastructure to the effectiveness of EIA and SEA studies, *Environmental Impact Assessment Review* Volume 25, Issue 2 , February 2005, Pages 123-142.

Virginia Tech, 2007. Form to Grade Peer Reviews. http://courses.cs.vt.edu/~cs3604/spring_2002/Writing/grading%20forms/WA5c%20form%20to%20grade%20peer%20review.htm as seen on February 14th 2007.

Walsh, E., Rooney, M., Appleby, L., Wilkinson, G., 2000. Open peer review: a randomized controlled trial. *The British Journal of Psychiatry*. 2000; 176: pp. 47-51.

Williamson, I., Rajabifard, A., Feeney, M.E., 2003. *Developing Spatial Data Infrastructures, From concept to reality*. Taylor & Francis, London, UK

Appendix A

Book

An Introduction to Spatial Data Infrastructures

REFeree REPORT GUIDELINES

1. Every chapter submitted is sent to a minimum number of three reviewers. It is therefore often necessary to receive all three reviews before an objective decision as to whether to publish can be made.
2. In conducting your review, please remember that the book *An Introduction to Spatial Data Infrastructures* is dedicated to the publication of original work about any aspect of Spatial Data Infrastructures and related areas. We encourage chapters which cover the main aspects each chapter topic described; Introduction (What is SDI?), Spatial Data, People, Standards, Technology, Policies, Cases, and Epilogue.
3. We discourage chapters strongly oriented to specific commercial products. On the other hand, we encourage chapters using clear examples explaining a specific chapter-related issue. Please take this into consideration when reviewing your paper.
4. In the rationale of your recommendation, please focus on WHY the chapter should be accepted/rejected, rather than summarising the chapter.
5. Please address the following in your motivations and comments to authors: originality, didactic level, technical quality, knowledge of area, presentation, and accessibility to the broad Spatial Data Infrastructure audience.
6. Try to be as informative as possible in your motivations and comments. In particular, if you are recommending acceptance, try to include comments on how to improve the final version of the chapter. If you feel the chapter should be rejected, try to describe the issues that must be addressed to make the paper acceptable.
7. You are welcome to submit your review in digital form. Save this page, edit it to insert your comments, and then return the completed form. Alternatively you can download the MS Word version and return the completed Word Document file. You are welcome to return reviews in digital form by email.
8. The review is an individual assignment and should be submitted before Friday 27 October 2006, 11.00 AM. This review forms 30% of the final mark of the course Spatial Data Infrastructures. At the last page, (page 8) you will find the criteria that will be used to assess your Book chapter review. Before Monday 25th of September you are in the position to choose one of the eight book chapters; 1) What is SDI?, 2) Spatial Data, 3) People, 4) Standards, 5) Technology, 6) Policies, 7) Cases, 8) Epilogue.

Please return your reviews by email to

Dr. ir. Joep Cromptvoets,
Centre for Geo-Information
Wageningen University
Droevendaalsesteeg 3
P.O. Box 47, 6700 AA, Wageningen
The Netherlands
Email: joep.cromptvoets@wur.nl
Tel. 0317-474399

In case you have any questions, do not hesitate to contact us.
We look forward to your reviews.

Arnold Bregt and Joep Cromptvoets

Name Reviewer:

Chapter Number:

Chapter Title:

Book Title: An Introduction to Spatial Data Infrastructures

1. Is the chapter written in grammatic and clear English? Yes / No
Motivate briefly your choice.

In general, the sentences are grammatically quite good. However, I get the idea that it has been made to finish before a deadline, because a lot of times, small words like ‘the an who’ can be put in to make it easier readable, while in other sentences, a lot of words are needed to explain something.
2. Is the writing style clear and appropriate to the work? Yes / No
Motivate briefly your choice.

The first part, ‘Introduction’ is quite weak, it is rather unclear how they want to explain and introduce the subject. Later on, the writing style becomes better. The parts ‘Spatial Data’ and ‘Infrastructure’ are excellent.
3. Are the title and subtitles appropriate to the content? Yes / No
Motivate briefly your choice.

It is strange that the ‘Introduction’ is actually one big example. What embarrasses me is that the following subtitle is called ‘Examples.’ ‘Spatial Data’ and ‘Infrastructure’ are excellent, as is the rest (Defining SDI, Driving forces for SDI development, conclusions, Further study)
4. Does the introduction of each chapter introduce well the content of the chapter? Yes / No
Motivate briefly your choice.

Introduction and ‘Examples’ are weak, the rest matches is appropriate
5. Is the length of the chapter appropriate to the content? Too short / Short / Appropriate/ Long / Too long
Motivate briefly your choice.

Yes it is, You get enough information out of it, and it doesn’t take a century to finish.
6. Are the conclusions borne out by the content? Yes / No / Not relevant
Motivate briefly your choice.

The conclusions match the content of this chapter
7. Are tables and figures necessary and appropriate? Yes / No / Not relevant
Motivate briefly your choice.

Not really; the first picture about water pollution is irrelevant. The first table with 'Key Issues' is quite unclear to me, it has no cohesion with the chapter. If you just didn't picture it, I wouldn't miss it.

Nevertheless, table 1.1, 1.2, 1.3 and 1.4 are of great importance, they are excellent.

The figure 1.2 with SDI components appears rare, if you made a block of all those blocks, it would've also been clear. What annoys me the more is that 'people' and 'technology' have their text upside down.

8. Is the list of references provided complete? Yes / No / Not relevant
Motivate briefly your choice.
- Yes, all the references mentioned in the chapter appear in the reference table.
9. Are the listed questions and exercises helpful? Yes / No / Not relevant
Motivate briefly your choice.
- Yes, they are relevant questions which makes you to look deeper into the text.

Please use the following codes in answering questions 10 - 15:

- 9 Outstanding (As good as the best course book chapters you have seen)
- 8 Strong Accept (Comparable to typical course book chapters published)
- 7 Accept (Solid chapter, but possibly not of international standard)
- 6 Weak Accept (Reasonable submission and worth support)
- 5 Neutral (Of borderline merit)
- 4 Weak Reject (I don't like it, but I won't object if others like it)
- 3 Reject (I would rather see this chapter not accepted)
- 2 Strong Reject (Must reject)
- 1 Unacceptable (I will be embarrassed to see this chapter in the book 'An Introduction to Spatial Data Infrastructures')

10. Originality [] i.e. novelty, degree of innovation

11. Didactic level [] i.e. designed to teach people something

12. Technical quality [] i.e. thoroughness, completeness

13. Relevance to Spatial Data Infrastructures [] i.e. interest to the Spatial Data Infrastructure community

14. Presentation [] i.e. clarity, readability

15. OVERALL RATING []

Motivate briefly your choice

16. Referee's confidence on the rating []
[L(OW)/A(VG)/H(IGH)]

Motivate briefly your choice

17. A short summary of the rationale for your recommendation (Maximum 5 lines):

Name Reviewer:

Chapter number:

Title Chapter:

Title book:

18. Detailed comments to authors (add additional pages (at least 2 A4-pages):

Additional PART ‘Non-assessment comments’

Name Reviewer:

**Book *An Introduction to Spatial Data Infrastructures*:
Title Chapter:**

Would you be willing to re-review this chapter if on the basis of your and other advice it is rejected on this first occasion? Yes / No

Signed:

Date:

Detailed comments for the Editor only (add additional pages as necessary) in order to express additional remarks (information):

Assessment criteria “SDI-book Chapter review”

This assignment is a completely individual one. During the last weeks of the course Spatial Data Infrastructures (GRS21306), you have to review a chapter of the book *An Introduction to Spatial Data Infrastructure* using the template as presented before. You have the opportunity to choose the chapter before Monday 25 September 2006. The review of the chapter has to be strongly linked to subject of the topic described.

Assessment criteria

- Individual activity
- Time of chapter decision (before 25 September 2006)
- Time of Review submission (before Friday 27 October 2006, 11.00 AM)
- Review relevant for subject of book chapter
- Quality of the motivations regarding questions 1 -16.
- Quality of the short summary (question 17)
- Quality of the detailed comments to authors (question 18)
- Size of the detailed comments (at least two pages A4)
- Style of writing (clear, understandable, logical structured)
- Originality of motivations and comments (not too much copy & paste)
- Thoroughness and Completeness of the review
- Use of references to existing literature (books, reviewed papers, internet pages) in order to justify the motivations and detailed comments

Appendix B

Revised version of:
Virginia Tech, 2007. Form to Grade Peer Reviews.

How well did peer reviewer address Content of proposal (15 points)

- good choice of issue as subject
- reflects knowledge of subject
- persuasiveness (e.g., arguments to motivate importance)
- specific action proposed
- length is in assigned range (deduction if word count not specified)
- not repetitive

How well did peer reviewer address Organization of proposal (15 points)

- order in which ideas are presented
- state problem, make clear what is sought as a solution
- acknowledgments and references

How well did peer reviewer address Structure of proposal (15 points)

- good logical flow
- cohesive sentences (single thought)

How well did peer reviewer address Style and Tone of proposal (10 points)

- professional and formal, rather than chatty and casual
- clarity: are the points made clearly and precisely to the reader

How well did peer reviewer address Mechanics of proposal (20 points)

- spelling and typos
- sentence structure
- grammar and punctuation
- proofread by (the names must be signed by the proofreaders in order to get credit for this item)

Overall comments from reviewer to author (25 points)

- overall quality of your (the reviewer) comments to the author
- how well your comments match those of the GTA
- professional tone of your comments
- how helpful your comments are in helping the author revise the paper

Style of Peer Review (no positive points but failure in this category can lead to possible significant deductions)

- constructive and helpful

- specific, not just vague and/or critical -- show specifics of what author should do
- supportive, not "slamming"
- helps author with revision

Deductions (lateness, not following instructions) **Total Points for Peer Review**
(100)

Appendix C

Time and results of different methods. (Second part on next page.)

Review	Time method 1	Result method 1	Reading time	Time method 2	Result method 2
1_1	0.49	NO	6.01	3.34	5,7
1_2	0.45	NO	2.15	1.43	5,0
1_3	0.38	YES	8.17	3.46	7,1
1_4	0.40	YES	5.09	1.49	7,9
2_1	0.20	YES	3.36	3.31	9,3
2_2	1.09	YES	2.10	2.31	5,7
2_3	0.27	YES	6.34	2.51	10,0
2_4	0.22	YES	9.51	2.57	8,6
2_5	0.25	YES	5.03	2.26	9,3
2_6	0.10	NO	0.10	0.45	3,6
3_1	0.25	YES	5.06	2.31	9,3
3_2	1.02	NO	7.50	2.59	8,6
3_3	0.30	NO	7.10	1.25	5,7
3_4	0.36	NO	4.35	2.51	5,7
4_1	0.19	YES	6.11	2.33	9,3
4_2	0.28	NO	3.15	2.08	5,7
4_3	0.20	YES	3.43	2.00	7,9
4_4	0.52	NO	6.36	1.35	5,0
4_5	0.55	NO	6.03	1.03	5,0
4_6	0.48	YES	8.04	2.25	7,9
5_1	0.40	NO	13.10	4.15	7,1
5_2	0.30	YES	9.10	2.10	6,4
5_3	0.23	YES	10.25	2.28	7,1
5_4	0.38	NO	9.26	1.38	5,7
5_5	0.51	NO	6.35	1.16	7,1
6_1	0.55	YES	6.03	3.03	10,0
6_2	0.24	NO	4.57	2.36	7,1
6_3	0.17	YES	3.50	2.28	9,3
6_4	0.06	YES	5.51	1.27	8,6

Time method 3	Result method 3	Time method 4	Result method 4	Grade	Result cal 3
6.11	5,6	11.14	1	6,6	6,5
3.06	5,0	3.12	0	7,1	5,8
7.17	6,3	17.02	3	8,7	7,3
2.24	5,6	0	0	5,4	6,5
7.50	6,3	3.42	1	6,8	7,3
4.53	4,4	2.11	2	6,0	5,1
5.36	9,4	20.38	11	8,9	10,9
6.03	8,1	24.04	10	9,1	9,4
3.41	8,8	6.45	5	8,7	10,2
0.57	1,9	0	0	4,0	2,2
5.33	6,9	7.50	0	6,8	8,0
7.07	7,5	21.08	3	7,8	8,7
4.29	3,8	0	0	6,4	4,4
5.10	5,6	9.24	3	7,2	6,5
4.50	8,1	6.35	4	8,9	9,4
3.31	5,6	7.20	5	5,0	6,5
4.10	7,5	11.59	10	9,4	8,7
4.30	5,0	11.12	3	7,0	5,8
2.16	5,0	5.39	3		
3.47	6,9	9.31	7	6,4	8,0
8.43	7,5	24.50	7	8,1	8,7
4.20	5,0	8.56	3	7,9	5,8
3.58	5,6	13.55	5	7,4	6,5
3.25	4,4	8.59	6	7,7	5,1
2.45	6,3	18.56	2	7,4	7,3
5.21	9,4	9.55	6	9,2	10,9
4.41	6,9	10.12	2	7,4	8,0
3.46	8,1	11.27	7	9,1	9,4
3.31	7,5	10.57	5	8,4	8,7