Vinicius Ferreira de Castro

A FRAMEWORK TO IDENTIFY, MEASURE AND BENCHMARK OPERATIONS MANAGEMENT BEST PRACTICES

Dissertação submetida ao Programa de Pós-Graduação de Engenharia de Produção da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Engenharia.
Orientador: Prof. Dr. Enzo Morosini Frazzon.

Florianópolis
2016

Castro, Vinicius Ferreira de
A FRAMEWORK TO IDENTIFY, MEASURE AND BENCHMARK OPERATIONS MANAGEMENT BEST PRACTICES/
Vinicius Ferreira de Castro ; orientador, Enzo Morosini Frazzon
Florianópolis, SC, 2016
114 p.
Dissertação (mestrado) – Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia de Produção.
Inclui referências
1. Engenharia de Produção. 2. Melhores práticas. 3. Benchmarking. I. Morosini Frazzon, Enzo. II. Universidade Federal de Santa Catarina. III. Programa de Pós-Graduação em Engenharia de Produção. III. Título.

Vinicius Ferreira de Castro

A FRAMEWORK TO IDENTIFY, MEASURE AND BENCHMARK OPERATIONS MANAGEMENT BEST PRACTICES

Esta Dissertação foi julgada adequada para obtenção do Título de Mestre em Engenharia e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção.

Florianópolis, 3 de março de 2016.

_____
Prof. Fernando Antônio Forcellini, Dr.
Coordenador do Curso

Banca Examinadora:

_____
Prof. Enzo Morosini Frazzon, Dr.
Orientador
Universidade Federal de Santa Catarina

_____
Prof. Dalton Francisco de Andrade, Dr.
Universidade Federal de Santa Catarina

_____
Prof. Antonio Cesar Bornia, Dr.
Universidade Federal de Santa Catarina

_____
Prof. Jovane Medina Azevedo, Dr.
Universidade do Estado de Santa Catarina

Este trabalho é dedicado à minha filha
e à família que nos cerca.

AGRADECIMENTOS

Agradeço à minha filha, Nina, que me lembra, com seu doce sorriso, a ser todos os dias a pessoa que eu gostaria que ela se torne.

Aos avós, Jajá, Moacir e Leninha, e à toda família que nos cerca, pelo apoio incondicional, incentivo, dedicação, paciência e infinito amor.

Ao meu pai e à Re, que estavam comigo no início desta caminhada, e que não puderam me ver chegar ao seu final, mas que me ajudaram todos os dias com as boas lembranças e lições de vida, a superar os desafios que surgiram.

Ao meu orientador, Prof. Enzo, por dividir comigo esses anos e me incentivar em seguir em frente nos bons e nos maus momentos.

Aos professores e colaboradores do PPGEP e do ProLogIS, que sem dúvida alguma inspiraram muito da pequena contribuição que eu pude fazer com essa singela peça.

Obrigado.

RESUMO

Este estudo teve como objetivo apresentar um método para identificar as melhores práticas entre um conjunto de empresas que é adaptável e estaticamente relevante. Este estudo abrangeu a literatura sobre a benchmarking de melhores práticas, aplicou a Teoria de Resposta ao Item (TRI) em um questionário de avaliação de melhores práticas, e propôs um método para avaliar as melhores práticas que levou em consideração as críticas mais relevantes sobre o tema. A análise foi realizada com um questionário com 46 itens que continha um banco de dados de 302 respostas. O modelo da TRI adotado foi o modelo logístico de três parâmetros. As respostas no banco de dados foram dicotomizadas devido à restrição de dados. Pelos erros apresentados na calibração dos itens, foi possível assumir que os parâmetros foram estimados com uma boa margem de segurança. O "Framework para Identificação de Melhores Práticas" surgiu a partir da análise dos estudos identificados na revisão da literatura. O teste do framework demonstrou as oportunidades de aplicação, contribuindo para a discussão sobre o uso de ferramentas estatísticas, especialmente a TRI, para a avaliação das melhores práticas e benchmarking.

Palavras-chave: Melhores práticas, Benchmarking, Teoria de Resposta ao Item.

ABSTRACT

This study aimed to present a method to identify best practices among a set of companies that is adaptable and statically relevant. This study covered the literature on benchmarking of best practices, executed a test application of Item Response Theory (IRT) on a best practices assessment questionnaire, and proposed a method to assess best practices that address most relevant critique on the topic. The analysis was carried with a best practices assessment questionnaire with 46 items that had a database of 302 responses. The IRT model adopted was the three-parameter logistic model. The answers in the database were dichotomised because of data restriction. By the errors presented in the calibration of the items, it was possible to assume that the parameters were successfully estimated with a good margin of certainty. The Best Practices Identification Framework emerged from the analysis of the studies identified in the literature review. The framework test demonstrated its applicability and opportunities, contributing to the discussion of the usage of statistical tools, especially the IRT, to the assessment of best practices and benchmarking purposes.

Keywords: Best practices, Benchmarking, Item Response Theory.

RESUMO EXPANDIDO

Introdução

Empresas se interessam por uma melhor maneira de executar seus processos por diversos motivos, como melhorar performance e mitigar riscos. Se uma empresa melhora sistematicamente seus processos, espera-se que ela reduza custos e defeitos, aumente a qualidade de seus produtos e a satisfação de seus clientes (DEMING, 1982; JURAN, 1999; SCHONBERGER, 1986).

Conceitualmente, melhores práticas se referem às práticas gerenciais e modelos que se mostraram bem sucedidas em empresas exemplares (LESEURE et al., 2004). Esse conceito se relaciona com o conceito de *benchmarking*, que é definido como sendo o processo pelo qual uma organização melhora sua performance por meio da comparação com de seus produtos, serviços e processos com outras empresas que são reconhecidas por apresentar melhores práticas (CAMP, 1989).

Desde Camp, o conceito de *benchmarking* foi aplicado com sucesso em praticamente todos os setores (DATTAKUMAR; JAGADEESH, 2003), contudo, muitos estudos neste tema foram criticados quanto à universalidade das melhores práticas que eles identificaram (LESEURE et al., 2004; SOUSA; VOSS, 2008; WELLSTEIN; KIESER, 2011), pois, uma melhor prática identificada em um estudo particular só poderia ser considerada universal se ela de fato estivesse correlacionada a uma melhor performance das empresas que a adotaram, o que exigiria amostra relevante e métodos estatísticos que não são comumente adotados nos estudos de *benchmarking*.

Uma das principais alternativas para lidar com essas críticas advém da aplicação da Análise Envoltória de Dados (*Data Envelopment Analysis* – DEA, em inglês) (AMADO; SANTOS; SEQUEIRA, 2013; DAI; KUOSMANEN, 2014; RUIZ; SEGURA; SIRVENT, 2015). Entretanto, essa abordagem, inerentemente quantitativa, também apresenta problemas, pois frequentemente está sujeita aos dados disponíveis, que raramente são suficientes para que seja possível analisar as melhores práticas. De fato, os estudos de DEA são eficientes em identificar as unidades com melhor desempenho, mas não justificam satisfatoriamente quais aspectos explicam esse melhor desempenho.

A Teoria de Resposta ao Item (*Item Response Theory* – IRT, em inglês) pode ser uma alternativa para compor um método de análise mais versátil, que permita mudanças na coleta de dados em estudos longitudinais, sem que se perca as referências de comparação dos traços

latentes (ALEXANDRE et al., 2002a; BALBIM JUNIOR; BORNIA, 2011; RASCH, 1993; STOCKING; LORD, 1983).

Dessa forma, este estudo propõe um método para identificar, medir e realizar o *benchmarking* de melhores práticas na gestão de operações. Dentre suas principais contribuições, destacam-se: os métodos adotados para revisão de literatura, que combinaram técnicas bibliométricas e de análise de conteúdo; a análise da literatura, que consolidou os principais métodos adotados na literatura para realizar o *benchmarking* de melhores práticas; a aplicação da IRT na análise de um questionário de *benchmarking* de melhores práticas respondido por 302 empresas; que serviu para validar o instrumento de pesquisa; o *framework* proposto para fazer o *benchmarking* de melhores práticas de gestão de operações.

Objetivos

O objetivo geral deste estudo é apresentar um método, que seja estatisticamente relevante e flexível, para identificar, medir e comparar melhores práticas em um grupo de empresas.

Os objetivos específicos são:

- Revisar a literatura sobre *benchmarking*;
- Identificar oportunidades para contribuição com os estudos mais recentes sobre *benchmarking* e melhores práticas;
- Analisar os itens de um questionário de *benchmarking* aplicando a IRT;
- Propor um *framework* que possa resolver alguns dos problemas metodológicos identificados na literatura.

Metodologia

Os passos gerais da metodologia de estudo consistiram em: (1) proposição de um *framework* a partir da análise da literatura; (2) teste do *framework* em um instrumento de *benchmarking*; (3) análise de resultados.

A revisão de literatura empregou dois softwares para análise bibliométrica: *Hammer Nails Project* (KNUTAS et al., 2015) e *VOSViewer* (VAN ECK; WALTMAN, 2010). Foram analisados 674 artigos da base do *Web of Science*, e os artigos mais relevantes foram identificados por três critérios de citação (citações totais, citações *in-degree* e *pagerank*) e pela análise de nuvens de cocitação.

Os artigos mais relevantes foram então analisados em profundidade, identificando-se os conceitos associados ao tema, os métodos adotados e as críticas mais relevantes aos estudos da área.

A aplicação da IRT analisou 302 respostas de indústrias brasileiras que responderam o questionário do IEL/SC (Instituto Euvaldo Lodi de Santa Catarina) entre 2010 e 2013. Apesar da amostra não ser representativa da indústria nacional ou catarinense, os dados foram suficientes para parametrizar os itens com erro satisfatório, o que foi feito usando o software *Bilog-MG* (ZIMOWSKI et al., 2003).

As respostas na base de dados (originalmente em uma escala *likert*) foram dicotomizadas e foi aplicado o modelo logístico de três parâmetros (ANDRADE; TAVARES; VALLE, 2000a), que representa a probabilidade P de uma companhia com uma maturidade θ ter implementada a prática *i*, conforme equação abaixo.

$$P_i(X = 1|\theta, a_i, b_i, c_i) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta - b_i)}}$$

Na equação, os parâmetros referem-se a: parâmetro *a* – capacidade discriminatória do item; parâmetro *b* – dificuldade do item; e parâmetro *c* – probabilidade de um item ser respondido casualmente de forma positiva.

Argumentou-se pela adoção do parâmetro c, pois, não é raro que um item seja interpretado erroneamente pelo respondente, ou ainda, que uma empresa adote uma boa prática de gestão sem que possua maturidade organizacional para tanto.

Os bons e os maus itens identificados foram então comparados entre si e analisados por um painel de especialistas, que apontaram os aspectos que possivelmente justifiquem a qualidade dos itens.

Resultados e discussão

A análise bibliométrica identificou o crescente interesse pelo tema mantendo-se a tendência de crescimento da produção de artigos iniciada em meados da década de 90, sendo que 47% dos artigos na base foram produzidos nos últimos 5 anos, e 2015 corresponde a mais de 10% da produção total.

Cooper, Sherman e Despotis foram os três autores mais relevantes de uma relação de mais de 20 autores relevantes identificados. Bem como foram identificados os periódicos mais citados.

A análise de citação identificou 29 artigos com ao menos uma citação *in-degree* e *pagerank* (PAGE et al., 1998) superior a 4,6e-05, além de outros 6 referências relevantes que não constavam na base analisada.

A análise de cocitação evidenciou 74 artigos organizados em seis clusters, que foram agrupados em dois grandes grupos com claras distinções entre si: um grupo de artigos apresentou uma abordagem mais gerencial do tema enquanto outro grupo focou mais em modelagem matemática para análise de eficiência e otimização.

Ao todo, 45 artigos foram analisados em profundidade. Compuseram a análise, além dos artigos mais relevantes de acordo com os critérios de citação e pelo mapa de cocitação, 6 artigos publicados nos últimos três anos nos cinco periódicos mais citados pela base analisada.

Referente à análise da IRT, curva de informação total de teste apontou avalia melhor empresas com maturidade entre -0,5 e 2,0, o que indica que o questionário, de modo geral, está difícil para as empresas avaliadas.

Os itens que avaliam a performance das organizações não se mostraram eficientes em diferenciar as empresas, ao passo que os itens que avaliam práticas foram considerados adequados.

Conclusão

O presente estudo cobriu a literatura de forma sistemática, evidenciando as oportunidades de contribuição acadêmica para o tema.

As críticas feitas aos estudos de *benchmarking* e aos estudos de DEA puderam ser endereçadas com o emprego da IRT, que permitiu a adoção de um instrumento de coleta flexível, que pode se adaptar com o tempo a novas práticas organizacionais permitindo ainda a comparação da maturidade de gestão das empresas. A aplicação teste do *framework* proposto sustenta sua aplicabilidade.

Como orientações para pesquisas futuras, sugere-se que, a partir de uma base de dados maior, possa ser possível aplicar variações da IRT para escalas *likert* bem como outros modelos da Teoria. Sugere-se também que os passos seis e nove do modelo proposto sejam testados juntamente com o IEL/SC ou em outro estudo de caso. Por fim, o passo 7 do *framework* foi realizado utilizando-se apenas uma análise de correlação dos itens de prática com a performance das empresas, o que pode ser aprofundado com outras técnicas estatísticas ou até mesmo com a aplicação do DEA.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

IEL/SC – Instituto Euvaldo Lodi de Santa Catarina
IRT – Item Response Theory
DEA – Data Envelopment Analysis
DMU – Decision Making Unit
UFSC – Universidade Federal de Santa Catarina

SUMÁRIO

1. INTRODUCTION

Since human kind has started to ask which is the best way to do the necessary tasks and activities, the discussion on best practices emerged and grew. Going as far as Sun Tzu, the ancient Chinese general that proposed *The Art of War*, passing through the discussions on benchmarking and the knowledge management practices, the important question has ever been "is there a better way to do what I do"?

A company may be interested in a better way of doing what it does for several reasons: improving performance, enhance competitiveness and mitigate risks are, perhaps, the most obvious ones. For instance, if a company improve systematically its process, it is expected that it will reduce costs and defects, and increase quality, consumer satisfaction and, ultimately, revenues (DEMING, 1982; JURAN, 1999; SCHONBERGER, 1986).

Conceptually, *best practices* refers to the "management practices and models that have been successful in exemplar firms" (LESEURE et al., 2004) and, accordingly to these authors, it can be traced back to the *World Class Manufacturing* work of Schonberger (1986).

This concept is closely related to the *benchmarking* concept, which is defined as a process that allows an organization to increase its performance by comparing its products, services and processes with others that are recognized as representing best practices (CAMP, 1989).

Since Camp, the concept of benchmarking has been applied successfully to almost all industries (DATTAKUMAR; JAGADEESH, 2003), evolving from the approach that focused mainly on the measurement of performance to that which focuses on the management activities and practices that lead to superior performance (VOSS; CHIESA; COUGHLAN, 1994).

However, a significant number of studies of benchmarking or best practices assessment has an important drawback regarding the universality of the best practices they have identified (LESEURE et al., 2004; SOUSA; VOSS, 2008; WELLSTEIN; KIESER, 2011). A best practice identified in a particular study can only be accepted as a universal best practice if it was applied in several companies and tested statically, if it is in fact correlated with a best performance of the companies that have adopted it in contrast to those that had not.

That kind of research limitation is not a novelty. Several studies in different knowledge fields have applied cross-validation models and similar so their findings could be generalized (FORKER; MENDEZ, 2001; LEE et al., 2005; ULUSOY; IKIZ, 2001). The problem with the adoption of this kind of methodology is the difficulty and prohibitive costs to collect enough data so the tests can be significant.

This issue is especially relevant when one is looking for best practices adopted by companies because they are usually emerged in a competitive environment, which causes them to constantly evolve their practices in order to sustain their competitive advantage. The consequence of those changes in the management practices adopted by companies is that the studies become obsolete very fast (KUULA; PUTKIRANTA; TOIVANEN, 2012; LAUGEN et al., 2005), and the data and the instruments applied to collect it cannot be exploited further more.

The main alternative for the critique on the benchmarking of best practices has been the adoption of Data Envelopment Analysis (DEA) to identify the benchmarks and the best practices (AMADO; SANTOS; SEQUEIRA, 2013; DAI; KUOSMANEN, 2014; RUIZ; SEGURA; SIRVENT, 2015).

The main drawback of these DEA studies is that they are always limited by the data available – to a best practice be identified it must have been assessed and data must be collected previously – one can never guarantee that the data available covers all the possible best practices there is. In order to do that, it would be necessary a flexible method, with a flexible questionnaire that can evolve by assessing new practices and discarding obsolete ones at the same time that it permits the comparison between units.

The Item Response Theory (IRT) may be an alternative for these drawback because it permits flexibility to change the questionnaire with new practices without losing its capability to compare units. One can remove and add new items without changing the final score of the respondents, allowing the comparison between them, even if they responded different questions.

The IRT has been applied to measure latent traits of individuals which cannot be observed directly (such as intelligence, knowledge of a certain subject, or companies management maturity) by defining the probability of an individual to give specific answers to the items of a questionnaire (ANDRADE; TAVARES; VALLE, 2000a; BALBIM JUNIOR; BORNIA, 2011; RASCH, 1993; STOCKING; LORD, 1983).

IRT is usually applied in areas such as education and health, and has few cases in management or operations research (CARROLL; PRIMO; RICHTER, 2014; MOREIRA JUNIOR, 2010; VASCONCELOS; LEZANA; ANDRADE, 2013). Noticing that Pacheco, Andrade, & Bornia (2015) published the first study using the IRT explicitly in a benchmarking context.

1.1. OBJECTIVES

From that reasoning derives the following question: how a method to identify best practices can be statistically relevant so it's findings can be generalized and at the same time be adaptable, changing gradually its instrument to adapt to the new practices that emerges without losing previous data?

1.1.1. General objective

Considering that questioning, the objective of this study is to present a method to identify best practices among a set of companies that is adaptable and statically relevant.

1.1.2. Specific objectives

Unfolded from the general objective, the specific objectives proposed are:

- To review the literature on benchmarking of best practices;
- To identify opportunities to contribute with the current research on benchmarking of best practices;
- To analyse the items of a benchmarking questionnaires using the IRT;
- To propose a framework that can address some of the issues identified in the literature review.

1.2. CONTRIBUTIONS OF THIS STUDY

In this study a questionnaire with 302 responses of small and medium sized companies was analysed with the Item Response Theory (IRT).

The adoption of IRT is justified because it is useful to understand the characteristics of the items of the questionnaire, such as difficulty and capacity to discriminate companies. In addition, it permits flexibility to change the questionnaire in the future, removing and adding new items without changing the final score of the respondents and allowing the comparison between them, even if they responded different questions. This flexibility turned out to be an important aspect of the proposed framework.

The novel deliverables of this study are:

- the methods adopted to review the literature, that used bibliometric analysis and cluster analysis software;
- the consolidation and analysis of the methods for benchmarking of best practices applied by other authors, presented in the literature review; and
- the use of IRT to analyse and validate a research instrument; and
- the flexible framework proposed for benchmarking of best practices.

1.3. LIMITATIONS

This study has some limitations regarding the literature research, the application of the IRT, and the testing of the framework.

The literature of benchmarking of best practices was reviewed systematically using some bibliometric procedures to endorse the most relevant articles that went through in-depth analysis. Although extensive, covering more than 600 articles, the review was limited (due to the software chosen) to the database of *Web of Science*. The literature on IRT and other relevant methods were also extensively reviewed even without a use of a systematic method. The literature review did not address more deeply each of the best practices analysed by the questionnaire itself.

The application of the IRT considered the three-parameter logistic model due to data available that were not sufficient to run a model more appropriate for *Likert* items.

Only some steps of the proposed framework were tested. To a complete test, it would be necessary to IEL or other consulting group to systematically use it for a certain period of time and collect enough data.

1.4. DOCUMENT STRUCTURE

This document is an article compendium that presents three articles elaborated during the research period, from 2013 to 2015. The articles are presented in the Table 1.

Table 1 – Articles in this compendium.

Article	Authors	Publication
Benchmarking of best practices: overview of scientific literature	Castro, Vinicius Ferreira de; Frazzon, Enzo Morosini	Submitted to *Benchmarking, an International Journal*
Analysis of a Questionnaire for Best Practice Assessment Using Item Response Theory	Castro, Vinicius Ferreira de; Frazzon, Enzo Morosini; Andrade, Dalton Francisco de; Bornia, Antonio Cesar	Submitted to the *European Journal of Operational Research*
A framework to identify and measure best practice adoption	Castro, Vinicius Ferreira de; Frazzon, Enzo Morosini	Published in the annals of *ciKi 2015 - Congresso Internacional de Conhecimento e Inovação*

Source: Author (2016).

The session 3 presents the first article that covers the literature review on benchmarking of best practices. The second article, that address the application of the IRT on a questionnaire designed to benchmarking of best practices is presented in session 0. Finally, the third article, in session 5, discuss the method proposed by this study, taking into consideration what was learned in the previous two studies.

As one may expect from articles that are part of the same study, the three articles – especially the third one, that was designed on top of the other two – have similarities and redundancies that were necessary to make a reader of a single article understand its context. In order to avoid these redundancies in this document, the articles were slightly adapted in terms of content, especially the introductions and reviews of literature. Other format adaptations were also necessary to fit the articles into the required standards of UFSC, that is different than the submitted journals and event.

The remaining of this document consists in the presentation of the methodological procedures, adopted in each of the three articles – in session 2, and the conclusion and future research suggestion – in session 0.

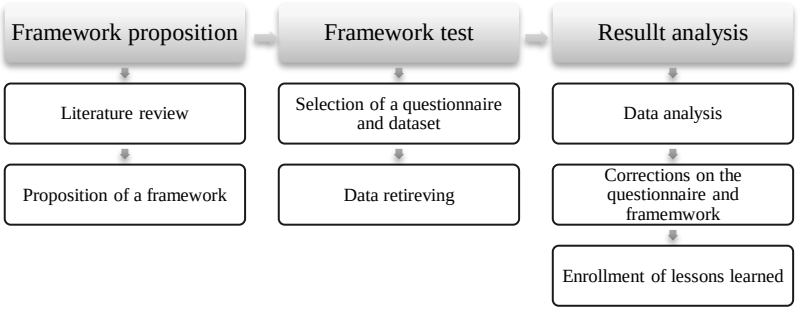
2. METHODOLOGICAL PROCEDURES

The procedures applied to propose and test the *Best Practices Identification Framework* started with a literature review of studies that identified and analysed best practices in business and operations management, highlighting the steps adopted by different authors.

The framework proposed emerged from the analysis of the studies identified in the literature review. To test the framework, it was selected a questionnaire to assess best practices that was already in use by IEL/SC. The questionnaire was already applied in 302 industrial companies and the data retrieved was analysed using the IRT.

The general steps adopted in this study are presented in the Figure 1.

Figure 1 – General steps adopted in the study development.



Source: Author (2016).

2.1. RESEARCH DESIGN

Regarding its nature, this research is classified as applied research, since it is concerned with the practical implications of the results (GIL, 2010).

Considering its objectives and procedures, it may be classified as exploratory, in its literature review phase, when were conducted a bibliographic research; and descriptive (GIL, 2010), when a case study (VOSS; TSIKRIKTSIS; FROHLICH, 2002) was conducted to verify the

practical application of the proposed framework. In some extend, one may consider this as a mixed methods procedures research (CRESWELL, 2009), since it combines quantitative and qualitative approaches.

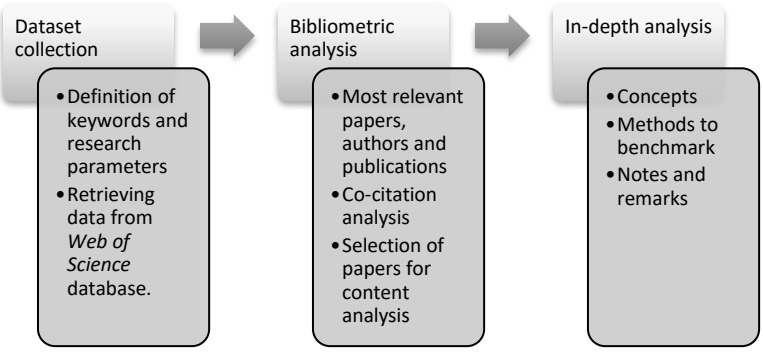
The research was designed as such to broaden understanding of the research context and build a better understanding of the results.

2.2. PROCEDURES FOR LITERATURE REVIEW

The first initial steps of the methodological procedures for literature review consisted in procuring a dataset from the Web of Science and run bibliometric analysis using two computer programs: *Hammer Nails Project* (KNUTAS et al., 2015) and *VOSViewer* (VAN ECK; WALTMAN, 2010). From the bibliometric analysis, the main papers in the field were identified, and its contents were analysed for the main concepts, methods, notes and remarks.

The Figure 2 presents the main steps of the methodological procedures.

Figure 2 – Main steps of methodological procedures for the literature review.



Source: Author (2016).

2.2.1. Dataset collection

The dataset was retrieved from *Web of Science* using the search parameters presented in the Table 2.

Table 2 – Search parameters.

Collection	Web of Science main collection
Topics	benchmarking AND best practices
Research areas	engineering OR business economics OR computer science OR operations research management science OR public administration
Document type	Article
Index	SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI

Source: Author (2016).

Although other databases were not considered for the bibliometric analysis due to software limitations (*Hammer Nails Project* software didn't run with *Scopus* or other databases), it was hoped that the *Web of Science* were sufficiently representative of scientific literature in management and operations engineering. This limitation was also addressed by incorporating for the in-depth analysis phase of the study new papers from relevant authors and publications that were not present in the dataset retrieved from *Web of Science*.

2.2.2. Bibliometric analysis procedures

The analysis identified the important authors and journals in the dataset based on the number of occurrences and citation counts. A citation network dataset retrieved from *Web of Science* were created and used to identify the important papers.

The main authors and publications were identified by two parameters: 1) number of articles in the dataset; and 2) citation by the articles in the dataset.

The most important papers and other sources were identified using three importance measures: 1) the in-degree citation network, 2) the citation count provided by *Web of Science*, and 3) the *PageRank* score (PAGE et al., 1998) in the citation network. The analysis also found often-cited references that were not included in the original dataset. Several authors analysed an article relevance by examining its backlink count, the number or articles that cites it, generally hypothesizing that more backlinks means higher importance (PAGE et al., 1998).

The co-citation analysis were done with the text mining software *VOSViewer* by creating a distance map (VAN ECK; WALTMAN, 2010)

with the references cited by the papers on the dataset collection. The analysis included only references that were co-cited at least five times, totalizing 74 articles. The distance map shows the references that are commonly cited together, revealing the main clusters of research and application of the theories and concepts regarding the dataset.

A pair of articles is considered to be "co-cited" when they both occur in the same reference list of a third article. Citation overlap between documents or frequent co-citation of two documents have proven to be strong indicators for document similarity (BOGERS et al., 2008).

2.2.3. In-depth analysis procedures

To summarize the main discussions, concepts and methods, this phase consisted in in-depth reading of 45 most relevant articles.

The articles selected for in-depth analysis consisted in:

- Most cited articles: the top 25 highest scoring papers in each citation criterion (in-degree citation, overall citation and PageRank) were identified using the measures separately. The results were then combined and duplicates removed.
- The most relevant articles in each cluster identified in the co-citation analysis. Duplicates with the most cited articles were excluded.
- Articles related to benchmarking best practices that were published in the ten most relevant publications in the last three years – from 2012 to 2015. The publications were identified as the ten most cited by the articles in the dataset.

Using these different selection criteria, it was hoped to address some of the drawbacks of considering citation as the only criterion for selecting relevant articles. By taking into consideration the clusters of the co-citation analysis we tried to avoid ignoring entire groups of articles that may represent new trends in the research field but have been less cited than the mainstream articles. By comprising novel articles, published in the last three years in relevant journals, we hoped to guarantee that new studies that have not been relevantly cited yet were also considered.

2.3. PROCEDURES FOR DATA ANALYSIS

The databased used for the analysis of the questionnaire consisted of 302 responses from different industrial sites of medium and small sizes from Brazil that participated on management development programs of IEL/SC between the years of 2010 and 2013.

The sample is not representative of industries of any size, region or sector, so one must be aware of the bias in the interpretation of the adoption of practices among industrials. To infer about how the practices are adopted and which one can bring more results, it will be necessary a better sampling. Yet, for the purpose of calibration of the items, which means finding the three parameters that characterize them, the sample is good enough (see Table 4).

The answers in the database were dichotomised considering that a company would have a practice implemented if it has answered 3 or above to the item related to that practice, and would not have it if the answer was 1 or 2. This procedure is similar to the one adopted by Alexandre *et al.* (2002), and is justified because of data restriction – to consider polyatomic items it would be necessary more data to estimate the item parameters with good degree of certainty.

Considering that, the IRT model applied was the three-parameter logistic model (ANDRADE; TAVARES; VALLE, 2000a) that represents the probability P of a company with a management maturity θ to have implemented the ith practice (see equation 1).

$$P_i(X = 1|\theta, a_i, b_i, c_i) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta - b_i)}} \quad (1)$$

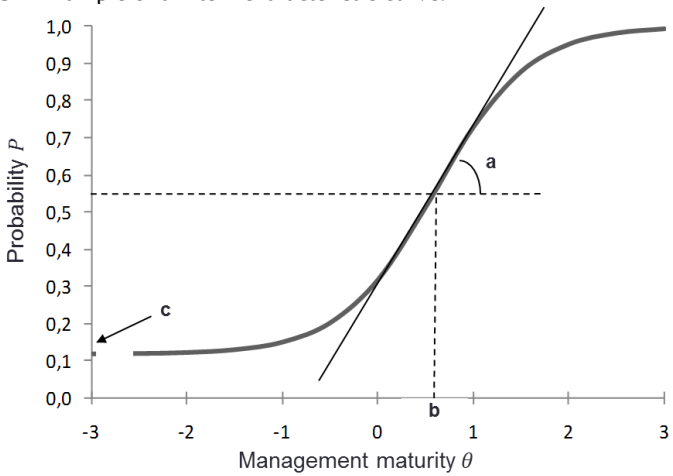
The parameters a, b and c of the ith item, shown in the Figure 3, have the following interpretation rationale:

- a is the capacity of the item to discriminate two companies with different management maturity. The higher is a, the higher is the variation of probability P of the company to have that practice adopted when company's maturity θ varies.
- b is the difficulty of the practice to be implemented, because the higher the b, the higher θ should be for that a

company present a reasonably good probability of having implemented the practice; and

- *c* is the probability of an item be answered positively even if the company has a very low θ.

Figure 3 – Example of an item characteristic curve.



Source: Andrade, Tavares and Valle (2000a).

In order to analyse the results and identify the best items in the questionnaire, the interpretation of the three parameters considered the reasoning described in the Table 3.

Table 3 – Reasoning to identify good and bad items

Parameter	Usual interpretation	Reasoning
a	Capacity of the item to discriminate respondents with different latent traits.	The more an item can discriminate respondents, the better it is.
b	Difficulty of the item.	It is important that the questionnaire contains items of different difficulties, so it can assess a wider range of latent traits.
c	Probability of the item to be answered correctly by a casual response.	A management practice can be implemented in a company that don't have corresponding management maturity. This parameter can measure this misalignment but also can measure the misunderstanding of the item by the respondent. In one way or another, the lower this parameter is, the better the item is, because it has more chance to effectively measure the latent trait assessed.

Source: Author (2016).

Considering this reasoning, the whole questionnaire was analysed following the steps presented in the Figure 4, so it was possible to identify and compare items and look for insights to understand what can cause an item to better discriminate companies and what can make it more difficult.

The comparison between items took into consideration items with similar difficulty but very different capacity to discriminate companies. A panel of experts that developed the original questionnaire and applied it in several companies discussed the insights on the aspects that make an item better.

The estimation of the parameters of the items was done using the software *Bilog-MG* (ZIMOWSKI et al., 2003) through a maximum likelihood estimation algorithm. The calibration has been achieved in 10 EM cycles with the largest change between cycles decreasing gradually until it was lower than 0.01, which indicates the algorithm converged smoothly.

The scale adopted in this study is the 0:1 scale, which means that the average score will be 0 (zero) and it will vary by 1 point for each standard deviation. This implies that, given a normal distribution of the sample, 95% of the assessments will be scored between 2 and -2.

The items that presented the largest error in the estimation of the parameter *b* are presented in the Table 4, noticing that these items were scored above 2 or below -2, so they were considered too difficult or too easy for the assessed companies.

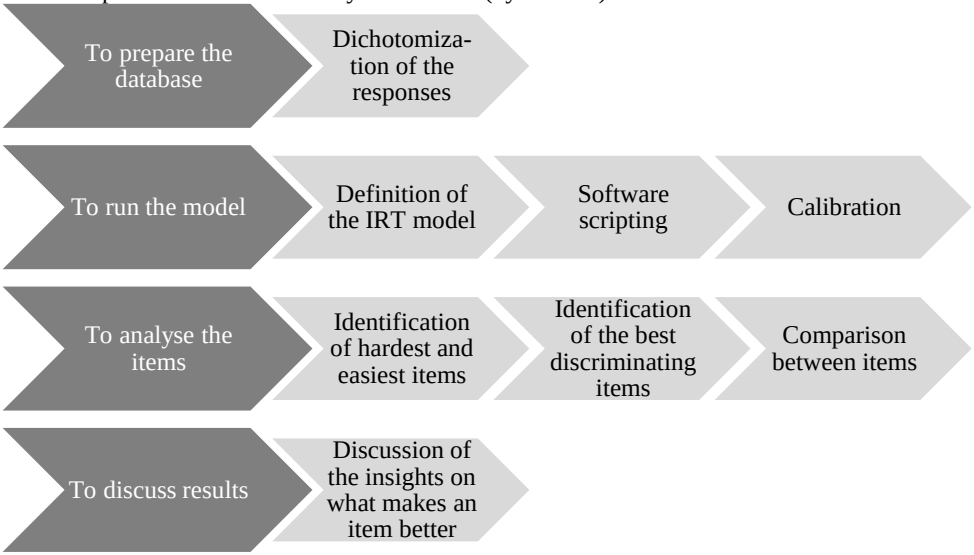
Table 4 – Highest standard errors on estimation of parameter b.

Item	Subject	Threshold (b)	Standard error
SA8	Injuries suffered by employees	-3,01234	0,93167
SA7	Absences due to illness	-2,13273	0,61209
PP24	New ideas proposed	2,75972	0,58875
SA3	Waste reduction	2,15679	0,4419

Source: Author (2016).

By the errors presented in the calibration of the items, it is possible to assume that the parameters were successfully estimated with a good margin of certainty.

Figure 4 – Steps of the method to analyse the items (by authors).



Source: Author (2016).

3. BENCHMARKING OF BEST PRACTICES: OVERVIEW OF SCIENTIFIC LITERATURE

The objective of this session is to overview the scientific literature on benchmarking of best practices, tracking the most important articles and understanding the similarity between studies. This study stablishes the reference ground for the next steps of the research.

From time to time, a review of the literature of a certain field of knowledge becomes important to identify and organize its main theories, new trends, most relevant works, influent authors and important publications.

Some studies previously reviewed the literature on benchmarking, such as Yasin (2002), Dattakumar and Jagadeesh (2003), Francis and Holloway (2007), Evans, Tisak and Williamson (2012) and Williams, Brown and Springer (2012).

Yasin (2002) researched electronic databases for published materials between the years of 1986 and 2000. They examined academic and practitioner literature and concluded that “despite the increasing scope of benchmarking activities and the number of organizations utilizing benchmarking, the field of benchmarking remains to large extend without an unified theory to guide its advancement” (YASIN, 2002, p. 234).

Dattakumar and Jagadeesh (2003) attempted a more comprehensive review of the literature on benchmarking, they organized and classified 382 publications, identifying the gaps for future research. The content classification proposed by the authors comprehends four groups: “general aspects and fundaments”, “specific applications and case studies”, “innovations/ extensions/ new approaches”, and “applicable to education sector”. They found that 170 publications belong to general aspects or fundamentals of benchmarking, 164 papers pertain to specific applications and case studies in benchmarking and 27 publications come under innovations/extensions/new approaches in benchmarking.

Francis and Holloway (2007) studied the literature on best practice benchmarking going through its concepts and typologies and focusing on criticisms of benchmarking, on how to evaluate its effectiveness, and on the notion of best practice.

Evans, Tisak and Williamson (2012) build up over Watson's (1993) work looking in the literature for changes in the five generations

of benchmarking proposed by him. They analysed 370 doctoral thesis and Master's dissertations from the years of 2003 to 2010 and concluded that none of the academic works had postulated a new generation of benchmarking.

Williams, Brown and Springer (2012) conducted a qualitative meta-analysis of 32 peer-reviewed sources from January 2005 to July 2010. Content analysis was applied to identify the reasons for reluctance for benchmarking and ways to overcome it. The study concluded that organizational leadership best practices could be applied to counter each of the major benchmarking reluctance concerns.

In general, the articles that reviewed the literature on benchmarking before explore the topic with different approaches and purposes. Yet, an important contribution that may be done to this kind of research (specifically for the benchmarking topic) consists in applying new and more rigorous bibliometric techniques to analyse the literature and draw quantitative and qualitative conclusions on the theme.

Although it has been pointed before that the literature on the this theme is wealth and diverse making it impracticable and undesirable to draw quantitative conclusions about the practice of benchmarking or the contribution of academic analysis of it (FRANCIS; HOLLOWAY, 2007), this study stands on the evolution of computational programs to propose a new quantitative and qualitative approach to the analysis of the literature of benchmarking of best practices.

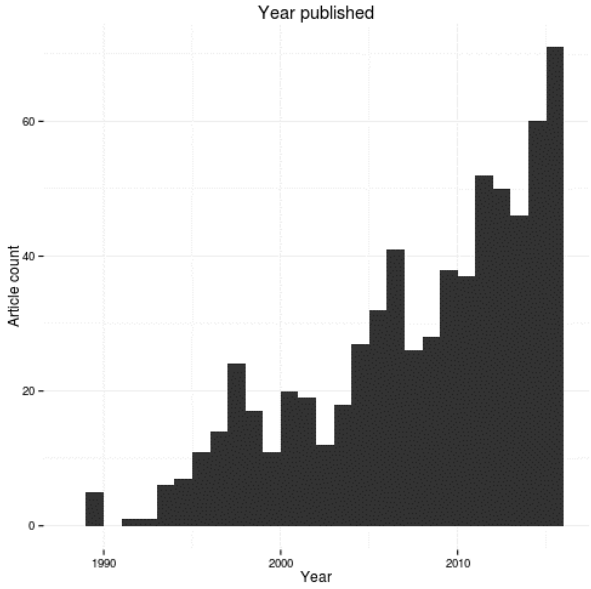
3.1. RESULTS OF THE BIBLIOMETRIC ANALYSIS

Looking for the most relevant papers, authors and publications, the bibliometric analysis in this study identified the relevance and the similarity between the 674 articles in the dataset (retrieved in November of 2015).

In order to do so, it was analysed the production frequency (categorized by year and by publication), the citation ranking (by overall and in-degree citation and by PageRank algorithm), and the co-citation map of references.

Starting with the production frequency, the Figure 5 shows the publication of articles per year.

Figure 5 – Articles published per year.



Source: Author (2016).

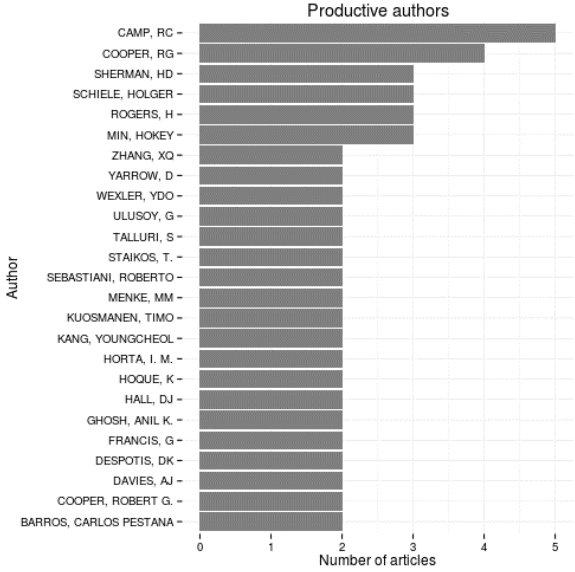
The first article published in the dataset is Camp (1989), who popularized the benchmarking of best practices concept by proposing a method to implement improvements in management practices by comparing itself with the best performers.

It is interesting to notice that the first decline in the literature output between the late 90s and early 2000s has been addressed by Dattakumar and Jagadeesh (2003), when they hypothesized that the topic was already mature.

Clearly the output assumed new peaks later showing a tendency of growing even today. The graph substantiates a consistent increasing in production of papers in this research topic. 47% of the articles in the dataset corresponds to the production of the last five years (between 2010 and 2015), and the year of 2015 alone represents more than 10% of all papers in the dataset.

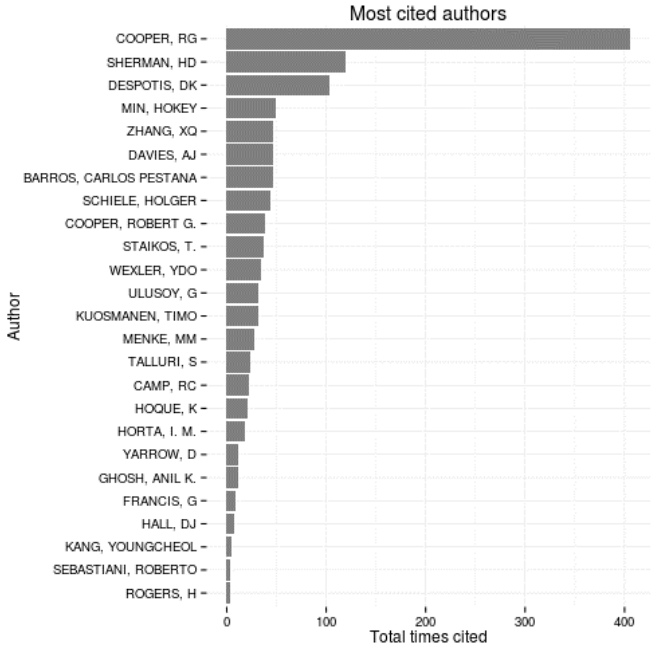
The main authors were identified by the amount of articles in the dataset (Figure 6) and citation count by the articles in the dataset (Figure 7).

Figure 6 – Authors with more articles in the dataset.



Source: Author (2016).

Figure 7 – Most cited authors by the articles in the dataset.



Source: Author (2016).

Among the main authors identified, one may notice that some have been researching best practices in different fields and with different methods even before the benchmarking seminal work of Camp (1989). That is the case of Cooper and Sherman. Although some of their first articles are not in the dataset (probably due to the parameters of the research), they certainly have become important influencers in this field.

Cooper's work is centred in the search of best practices of new product development (COOPER, 1979; COOPER; EDGETT; KLEINSCHMIDT, 1999, 2004a).

Sherman aims on application of Data Envelopment Analysis (DEA) in different sectors, such as banks (CHILINGERIAN; DAVID SHERMAN, 1996; SHERMAN; LADINO, 1995) and hospitals (CHILINGERIAN; DAVID SHERMAN, 1996; SHERMAN, 1984).

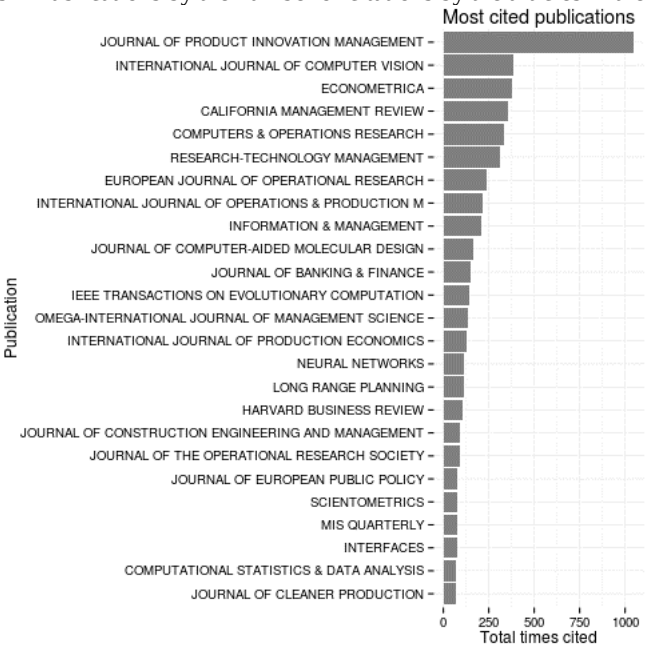
Also using DEA alongside with other methods, Despotis has been studying policies' efficiency for human development (DESPOTIS, 2005; SISKOS; DESPOTIS; GHEDIRI, 1994) and methodological issues of DEA application (DESPOTIS; SMIRLIS, 2002; SMIRLIS; MARAGOS; DESPOTIS, 2006).

The most important publications were identified using two criteria: the number of articles in the dataset and number of citations by the articles in the dataset. The Figure 8 and Figure 9 present results of the analysis of these criteria.

Figure 8 – Publications by the number of articles in the dataset.

Source: Author (2016).

Figure 9 – Publications by the number of citations by the articles in the dataset.



Source: Author (2016).

The most relevant articles were identified by three rankings: in-degree citation, overall citation (provided by *Web of Science)*, and PageRank score. The top 25 highest scoring papers were identified using these measures separately. The Table 5 presents the 29 articles that had at least one in-degree citation.

Table 5 – Most relevant articles identified.

Article	**In-degree**	**Overall citation (*Web of Science*)**	**PageRank**
MANAGING BANK PRODUCTIVITY USING DATA ENVELOPMENT ANALYSIS (DEA) (SHERMAN; LADINO, 1995).	9	73	5.92e-05
MANUFACTURING BEST PRACTICE AND PERFORMANCE	5	29	4.84e-05

Article	In-degree	Overall citation (*Web of Science*)	PageRank
STUDIES: A CRITIQUE (DAVIES; KOCHHAR, 2002).			
BEST MANUFACTURING PRACTICES - WHAT DO THE BEST-PERFORMING COMPANIES DO? (LAUGEN et al., 2005).	4	55	4.97e-05
WHAT CAN BENCHMARKING OFFER THE OPEN METHOD OF CO-ORDINATION? (ARROWSMITH; SISSON; MARGINSON, 2004).	4	35	5.02e-05
BENCHMARKING LOGISTICS PERFORMANCE WITH AN APPLICATION OF THE ANALYTIC HIERARCHY PROCESS (KORPELA; TUOMINEN, 1996).	3	37	4.87e-05
SERVICE QUALITY AND BENCHMARKING THE PERFORMANCE OF MUNICIPAL SERVICES (FOLZ, 2001).	3	31	4.76e-05
BENCHMARKING INITIATIVES IN THE CONSTRUCTION INDUSTRY: LESSONS LEARNED AND IMPROVEMENT OPPORTUNITIES (COSTA et al., 2007).	3	23	4.81e-05
BENCHMARKING - PERFORMANCE IMPROVEMENT TOWARD COMPETITIVE ADVANTAGE (LEMA; PRICE, 1995).	3	16	5.03e-05
PROCESS OF BENCHMARKING - A STUDY FROM THE AUTOMOTIVE INDUSTRY (DELBRIDGE; LOWE; OLIVER, 1995).	3	11	4.95e-05
BENCHMARKING TO IMPROVE THE STRATEGIC PLANNING PROCESS IN THE HOTEL SECTOR (PHILLIPS; APPIAH-ADU, 1998).	3	8	4.93e-05

Article	In-degree	Overall citation (*Web of Science*)	PageRank
BENCHMARKING BEST MANUFACTURING PRACTICES - A STUDY INTO FOUR SECTORS OF TURKISH INDUSTRY (ULUSOY; IKIZ, 2001).	3	6	4.80e-05
MADE IN THE 21ST CENTURY: HOW FAR HAVE WE COME ON THE JOURNEY TO EXCELLENCE? (YARROW; HANSON; ROBSON, 2004).	3	6	4.80e-05
ENHANCING THE EFFECTIVENESS OF BENCHMARKING IN MANUFACTURING ORGANIZATIONS (KUMAR; CHANDRA, 2001).	3	5	4.72e-05
NEW PRODUCT PORTFOLIO MANAGEMENT: PRACTICES AND PERFORMANCE (COOPER; EDGETT; KLEINSCHMIDT, 1999).	2	181	5.00e-05
FROM KNOWLEDGE TO ACTION: THE IMPACT OF BENCHMARKING ON ORGANIZATIONAL PERFORMANCE (DREW, 1997).	2	59	4.71e-05
PERSPECTIVE: ESTABLISHING AN NPD BEST PRACTICES FRAMEWORK (KAHN; BARCZAK; MOSS, 2006).	2	48	4.77e-05
MEASURING HUMAN DEVELOPMENT VIA DATA ENVELOPMENT ANALYSIS: THE CASE OF ASIA AND THE PACIFIC (DESPOTIS, 2005).	2	43	4.80e-05
AN ANALYTICAL METHOD FOR BENCHMARKING BEST PEER SUPPLIERS (FORKER; MENDEZ, 2001).	2	36	4.67e-05
DATA ENVELOPMENT ANALYSIS: PRIOR TO CHOOSING	2	28	4.72e-05

Article	In-degree	Overall citation (*Web of Science*)	PageRank
A MODEL (COOK; TONE; ZHU, 2014).			
EVALUATION OF PERFORMANCE OF EUROPEAN CITIES WITH THE AIM TO PROMOTE QUALITY OF LIFE IMPROVEMENTS (MORAIS; CAMANHO, 2011).	2	24	4.76e-05
PERFORMANCE ASSESSMENT OF CONSTRUCTION COMPANIES: A STUDY OF FACTORS PROMOTING FINANCIAL SOUNDNESS AND INNOVATION IN THE INDUSTRY (HORTA; CAMANHO; MOREIRA DA COSTA, 2012).	2	16	4.69e-05
BEST-PRACTICE BENCHMARKING USING CLUSTERING METHODS: APPLICATION TO ENERGY REGULATION (DAI; KUOSMANEN, 2014).	2	9	4.72e-05
THE ROLE OF CREATIVITY WITHIN BEST PRACTICE MANUFACTURING (HALL, 1996).	2	4	4.96e-05
SUCCESS FACTORS FOR INTEGRATING SUPPLIERS INTO NEW PRODUCT DEVELOPMENT (RAGATZ; HANDFIELD; SCANNELL, 2003).	1	262	4.60e-05
CONSORTIUM BENCHMARKING: COLLABORATIVE ACADEMIC-PRACTITIONER CASE STUDY RESEARCH (SCHIELE; KRUMMAKER, 2011).	1	10	4.80e-05
DO THE BEST NEW PRODUCT DEVELOPMENT PRACTICES OF US COMPANIES MATTER IN HONG KONG? (OZER; CHEN, 2006).	1	7	4.75e-05

Article	In-degree	Overall citation (*Web of Science*)	PageRank
THE FALLACY OF UNIVERSAL BEST PRACTICES (HARRINGTON, 2004).	1	4	4.77e-05
TRADING BEST PRACTICES-A GOOD PRACTICE? (WELLSTEIN; KIESER, 2011).	1	3	4.77e-05
ECONOMIC VALUE OF COMBINED BEST PRACTICE USE (LEE et al., 2005).	1	2	4.77e-05

Source: Author (2016).

Complementary to the articles in Table 5, six articles that were not in the dataset but were frequently cited were considered to be related to the research theme. These are presented in Table 6.

Table 6 - Other relevant articles not included in original dataset.

Article	In-degree	PageRank
FIRM RESOURCES AND SUSTAINED COMPETITIVE ADVANTAGE (BARNEY, 1991).	12	5,59E-05
BENCHMARKING BEST NPD PRACTICES-II (COOPER; EDGETT; KLEINSCHMIDT, 2004b).	9	6,06E-05
BENCHMARKING BEST NPD PRACTICES-I (COOPER; EDGETT; KLEINSCHMIDT, 2004a).	9	6,06E-05
THE COMPETITIVENESS OF EUROPEAN MANUFACTURING - A FOUR COUNTRY STUDY (VOSS et al., 1995).	8	5,65E-05
PDMA RESEARCH ON NEW PRODUCT DEVELOPMENT PRACTICES: UPDATING TRENDS AND BENCHMARKING BEST PRACTICES (GRIFFIN, 1997).	8	5,53E-05

BUILDING THEORIES FROM CASE STUDY RESEARCH (EISENHARDT, 1989).	8	5,12E-05

Source: Author (2016).

Two articles were excluded from the set because the authors considered they were not related to the research theme. The inclusion of *computer science* area in the search parameters may have caused noise in the results. Some articles of this area, focused on algorithm optimization, use the word "benchmarking" referring to the verification of the algorithm's efficiency, with no relevant contribution to any benchmarking theory.

The The main cluster (in red), with 24 references, is centred in Camp (1989) and has other relevant works in its core (BARNEY, 1991; CAMP, 1989; COHEN; LEVINTHAL, 1990; COOPER; EDGETT; KLEINSCHMIDT, 2004a; EISENHARDT, 1989; HAMMER, 1993; SPENDOLINI, 1992; SZULANSKI, 1996; WOMACK; JONES; ROOS, 1990) as shown in Figure 11.

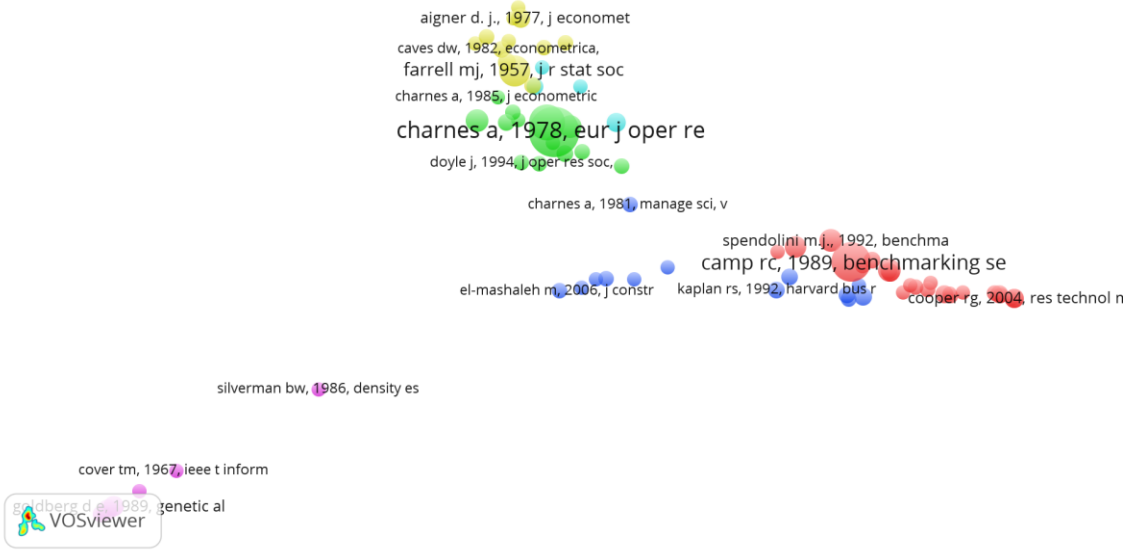
Figure 10 shows the overview map of the co-citation analysis. The distance map shows the references that are used to be cited together. A pair of articles is co-cited when they both occur in the same reference list of a third article.

In the map, each reference is indicated by a circle, which size indicates its relevance in terms of how many times it was cited with others. The distance between circles indicates the proximity between references in terms of co-citation (the closest they are, the more they were co-cited), and the colours indicates the different clusters identified.

In this analysis, it was considered all references that were co-cited at least five times, totalizing 74 articles. The *VOSViewer* software identified six clusters.

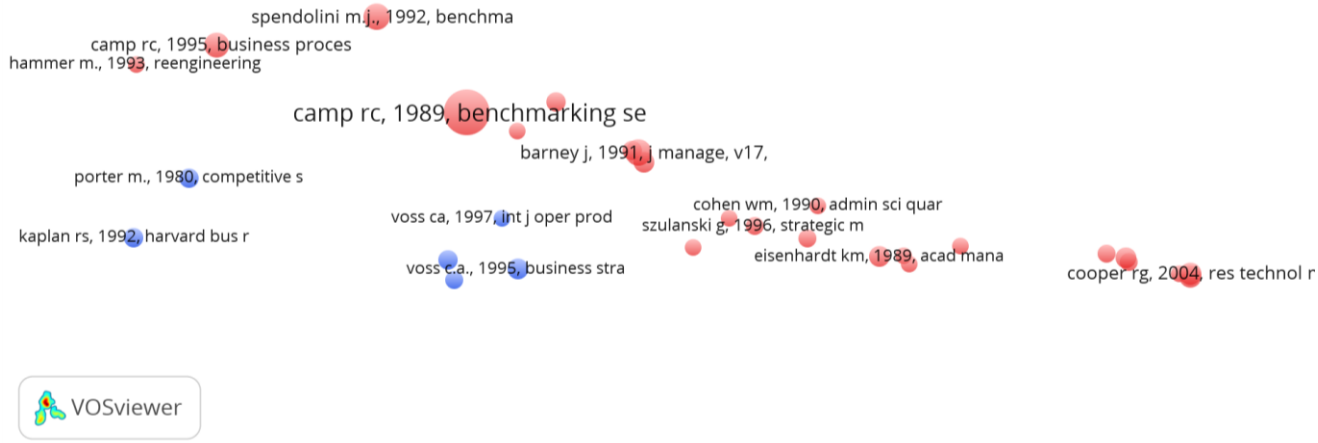
The main cluster (in red), with 24 references, is centred in Camp (1989) and has other relevant works in its core (BARNEY, 1991; CAMP, 1989; COHEN; LEVINTHAL, 1990; COOPER; EDGETT; KLEINSCHMIDT, 2004a; EISENHARDT, 1989; HAMMER, 1993; SPENDOLINI, 1992; SZULANSKI, 1996; WOMACK; JONES; ROOS, 1990) as shown in Figure 11.

Figure 10 – Co-citation map of references of articles of the dataset.



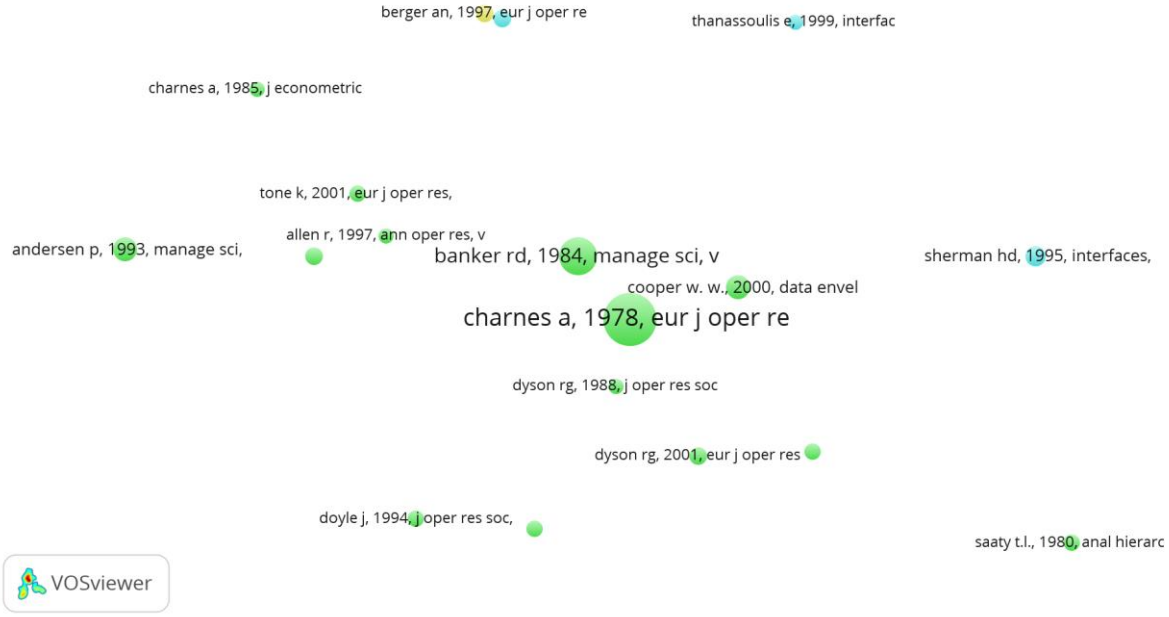
Source: Author in *VOSViewer* software (2016).

Figure 11 – Main cluster (in red) of the co-citation map.



Source: Author in VOSViewer software (2016).

Figure 12 – Second most relevant cluster (in green) of the co-citation map.



Source: Author in *VOSViewer* software (2016).

The second most relevant cluster (Figure 12 – in green) has 14 references and is centred in Charnes' applications of DEA and other related works (ALLEN et al., 1997; ANDERSEN; PETERSEN, 1993; BANKER; CHARNES; COOPER, 1984; CHARNES; COOPER; RHODES, 1978; DOYLE; GREEN, 1994; DYSON et al., 2001; DYSON; THANASSOULIS, 1988; SAATY, 1980; TONE, 2001).

Not so relevant in terms of co-citation counts, the other four clusters are:

- Cluster three (blue): it is related mainly to best practices in construction (EGAN, 1998; EL-MASHALEH; EDWARD MINCHIN JR; O'BRIEN, 2007; LEE; THOMAS; TUCKER, 2005). This cluster has some references that overlaps cluster one (HANSON; VOSS, 1995; HAYES; WHEELWRIGHT, 1984; PORTER, 1980; VOSS et al., 1995; VOSS; AHLSTROM; BLACKMON, 1997), which may reveal its relation to a more managerial approach of the theme.
- Cluster four (yellow): it is highly related to the cluster two, it also uses DEA to analyse efficiency (AIGNER; CAIN, 1977; COELLI et al., 1998; MEEUSEN; DEN BROECK, 1977).
- Cluster five (pink): related to computer algorithms optimization (GAREY; JOHNSON, 1979; GOLDBERG; HOLLAND, 1988; KIRKPATRICK et al., 1983), this cluster has seven references and is the most distant in the map. It was considered that the articles in this cluster were unrelated to the topic of this review. These references appear perhaps because the area of *computer science* was included in the search parameters.
- Cluster six (light blue): with only four references, this cluster, as the cluster four, is highly related to cluster two, with articles that presents more DEA applications (BOUSSOFIANE; DYSON; THANASSOULIS, 1991; DYSON; THANASSOULIS, 1988; SHERMAN; LADINO, 1995; TONGZON, 2001).

From the results of the co-citation map analysis, it became evident that there are two different major areas of studies of benchmarking of best practices, one that approaches the theme from the managerial perspective

(clusters one and three) and other that relies on mathematical modelling to identify the most efficient units in terms optimization of inputs and outputs, applying mainly the DEA theory (clusters two, four and six). The Table 7 presents the main differences of these two areas.

The overall result of the bibliometric analysis seems to be well balanced in terms of novelty and relevance of the identified articles. The citation analysis had a varied outcome, with articles ranging from 1989 to 2014 (see Table 8) and with more than 17% of them from earlier than 2010. In the other hand, the relevant papers identified in the two main clusters of the co-citation map tended to be older, with about 90% of them older than 2000. Both analysis has shown to be complementary – about 50% of the relevant articles identified in the co-citation map were also identified in the citation analysis.

This difference in results may be explained by the use of a mixed criterion to select relevant articles, including the use of PageRank algorithm, while the co-citation analysis tends to be more influenced by the antiquity of the article (the older it is, the higher the probability of being co-cited).

To complement the analysis, this research covered the issues of the top five publications (excluded the computer science specific journals) identified in the Figure 9 (presented previously) published in the last three years – from 2012 to 2015 – for articles related to the theme of research. This effort resulted in six articles (AMADO; SANTOS; SEQUEIRA, 2013; KUULA; PUTKIRANTA; TOIVANEN, 2012; MENKE, 2013; RUIZ; SEGURA; SIRVENT, 2015; SUAREZ; CALVO-MORA; ROLDÁN, 2016; WANG, 2013) that was separated for the next step of analysis – the final collection for in-depth analysis counted with 45 articles of which 12 (26,6%) were published in the past five years.

Table 7 – Major areas of research related to benchmarking best practices.

Perspective	Managerial	Efficiency modelling and optimization
Relevant authors	CAMP, 1989; COOPER; EDGETT; KLEINSCHMIDT, 2004a; EISENHARDT, 1989; SZULANSKI, 1996; VOSS; AHLSTROM; BLACKMON, 1997; VOSS, 1995	BERGER; HUMPHREY, 1997; CAVES; CHRISTENSEN; DIEWERT, 1982; CHARNES; COOPER; RHODES, 1978; CHARNES et al., 1985; DYSON et al., 2001; FARREL, 1957; TONE, 2001
Approach	Benchmarking as a process of comparing and implementing best practices that will support better performance. These studies look for best practices trying to understand how to implement them in different areas and companies. There are also studies that discuss concepts, methods and implications of benchmarking.	Benchmarking as a process of identifying the most efficient in a set of units, with high emphasis in mathematical modelling and definition of inputs and outputs. The main concern in these studies is to select good models, to identify efficient production units, to determine which production factors better differentiate these units, and which units are better benchmarks for others.
Related concepts	Total Quality Management, World Class Manufacturing, Business Strategy, Competition.	Data Envelopment Analysis (DEA), Production efficiency, Production function.

Source: Author (2016).

Table 8 – Analysis of the relevant articles identified in the bibliometric analysis.

Year published	Relevant articles by the citation analysis		Relevant articles by the co-citation analysis		Total relevant articles		Dataset collect-ion	
2015 - 2010	6	17,1%	0	0,0%	6	13,3%	316	47,0%
2009 -2000	16	45,7%	1	10,0%	17	37,8%	261	38,8%
Before 2000	13	37,1%	9	90,0%	22	48,9%	97	14,1%

Source: Author (2016).

Considering the points exposed in this section, the subset of articles selected to be analysed with more profundity in the next step of the research is composed of the 35 articles presented previously (Table 5 and Table 6), completed with other four articles identified in co-citation map (ALLEN et al., 1997; CAMP, 1989; COHEN; LEVINTHAL, 1990; SZULANSKI, 1996) and six articles selected from relevant journals that have been published in less than three years.

3.2. CONCEPTS, METHODS AND CONSIDERATIONS FROM LITERATURE

A total of 45 articles were selected to a deeper analysis. The analysis is presented in this session covering three aspects: 1. Concepts; 2. Methods for benchmarking; and 3. Notes and remarks.

3.2.1. Concepts

There are three important concepts that cover the most of the articles selected for in-depth analysis: *Benchmarking, Best practice and Data Envelopment Analysis*. These concepts are explored as follows.

3.2.1.1. Benchmarking

The early studies on benchmarking have seen it as practice that emerged from the context of total quality management system (LEMA; PRICE, 1995; VOSS et al., 1995) and Japanese philosophy – from the

dantotsu concept, meaning: striving for the best of the best (CAMP, 1989).

Several authors (CAMP, 1989; DAVIES; KOCHHAR, 2002; DELBRIDGE; LOWE; OLIVER, 1995) have presented and discussed two slightly different concepts of benchmarking:

- "as the continuous process of measuring products, services and practices against the toughest competitors or industry leaders"; and
- "as a continuous search for and application of significantly better practices that lead to superior competitive performance".

Benchmarking is frequently resumed as "the systematic search for best practices that lead to superior performance" (DREW, 1997; KORPELA; TUOMINEN, 1996; LEMA; PRICE, 1995; PHILLIPS; APPIAH-ADU, 1998; SZULANSKI, 1996), while *benchmark* refers to "a measured best-in-class achievement, a reference or measurement standard for comparison" (LEMA; PRICE, 1995, p. 30).

Although different, these concepts are not completely disconnected. They can even be considered as the same concept observed with different management maturity glasses: first a company begins to compare its performance with competition – performance benchmarking, then it will look for process proficiency and finally it will incorporate what it sees as best practice (KAHN; BARCZAK; MOSS, 2006).

Benchmarking is commonly classified by its focus on: *product, function or process, best practice or strategy* (DREW, 1997; KUMAR; CHANDRA, 2001)[1], and by the type of partner to benchmark with: *internal, direct competitor, related industry/ functional benchmarking, or unrelated industry/ generic benchmarking* (DREW, 1997; KORPELA; TUOMINEN, 1996). The most of the methods identified in this in-depth analysis are best practice benchmarking with direct competitor or with related industry and functions, except for Szulanski (1996) who discuss a method for best practice internal benchmarking.

The link between innovation and benchmarking in several studies substantiates that benchmarking is not about copying and imitating the

[1] This classification is quite similar to the one proposed by Watson (1993).

best performers, but rather improving upon their acknowledged best practices. Cohen and Levinthal (1990) argue that "the ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends is critical to its innovative capabilities" (p. 128). Costa et al. (2007) advocate that the benchmarking has an important role on introducing new ideas and innovations on organizations, on helping to accelerate and manage organizational changes and on creating a culture of continuous improvement. Although not directly linked, another study that also advocates the importance of creativity in the search for best practices is Hall (1996):

> "Whilst best practice manufacturing may define a set of required outcomes, the transformation process does not necessarily exist within the organization to deliver the level of performance. The attitude of management must be developed in order to open the mind, to entertain the concept of creativity, and then to innovate" (HALL, 1996, p. 120).

Moreover, the first studies on benchmarking have shown some concern on ethical aspects of imitating competition. The concept brought by Drew shows that very clearly: benchmarking is "the art of finding out, in a perfectly legal and aboveboard way, how others do something better than you do – so you can imitate – and perhaps improve upon – their techniques" (DREW, 1997, p. 427).

3.2.1.2. Best practices

An early definition of best practice associates it with "superior performance within an activity, regardless of industry, leadership, management, or operational approaches, or methods that lead to exceptional performance" (LEMA; PRICE, 1995, p. 30).

Generally it is accepted as the practices adopted by best performing companies, and they are frequently accepted as "best practices for all companies", not considering potential influential factors on the practice fit (LAUGEN et al., 2005). It is also generally accepted that the best practices leads to a superior performance (DAVIES; KOCHHAR, 2002; LEE et al., 2005)

A practice refers to the processes that a company has implemented to improve the way it runs its business (VOSS et al., 1995). The basic

elements of a practice are the routines – as pattern of behaviour that is followed repeatedly – that frequently are stated by organizational rules and standard operating procedures (WELLSTEIN; KIESER, 2011). "Every company needs to make strategic choices on where and how to compete (…), and in order to succeed, they need a set of practices" (KUULA; PUTKIRANTA; TOIVANEN, 2012, p. 108).

The term "best practices" has been suffering several criticisms in academic studies. Davies and Kochhar (2002) call attention to the use of the term "best" practice, that can get different meanings depending on the company and on the situation in which the practice is adopted and Harrington (2004) points out that a best practice may be very dependent on companies size and maturity[2]. Wellstein and Kieser (2011) argue that best practices are mere marketing constructs of management consultants and suggest reframing discussions on them around organizational routines and rules. More on this will be discussed in session 3.2.3 Notes and remarks.

3.2.1.3. Data Envelopment Analysis

Data envelopment analysis (DEA) is a mathematical programming based approach for measuring relative efficiency of decision making units (DMUs) that have multiple inputs and outputs (Charnes et al. 1978 APUD Cook et al. 2014). "DEA derives a single summary measure of efficiency for each DMU, which is based on the comparison with other DMUs in the sample" (HORTA; CAMANHO; MOREIRA DA COSTA, 2012, p. 85). Thus, to identify the best practices in a set of units, DEA identify the ones that uses the least resources to provide its volume of products (SHERMAN; LADINO, 1995).

DEA is mainly concerned with the estimation of efficiency of the DMUs, applying input-output weights that maximize the efficiency score of the evaluated units, while the benchmarks provided by DEA can be seen as a side-product of the envelopment problem (DAI; KUOSMANEN, 2014). In the circumstance of benchmarking, the efficient DMUs may not be necessarily a "production frontier", but rather a "best-practice" frontier – and in those cases, an important issue that must

[2] See Sousa and Voss (2008) for further discussion on context-dependent best practices.

be addressed is how to classify performance measures into inputs and outputs (COOK; TONE; ZHU, 2014).

DEA can also be used to assess productivity changes over time through the calculation of Malmquist Productivity Indices and their components, which is calculated considering the change in the efficiency rate of a particular DMU from one period to another and the change in the efficiency frontier from one period to another (AMADO; SANTOS; SEQUEIRA, 2013).

3.2.2. Methods used in relevant studies

The Table 9 summarizes the most systematized methods adopted by these authors to assess and identify best practices.

Table 9 – Methods used to assess and identify best practices.

Author(s)	General steps of the method for benchmarking	Main characteristics
(CAMP, 1989; DREW, 1997; KORPELA; TUOMINEN, 1996; SCHIELE; KRUMMAKER, 2011; SZULANSKI, 1996)	1. To define a research focus; 2. To plan research agenda and interview guideline; 3. To select target benchmarking partners; 4. To systematically collect and analyse data; 5. To identify and propose generalizable best practices; 6. To plan and implement improvements to increase performance.	• Case study approach; • Qualitative data focus.

(FORKER; MENDEZ, 2001; KUULA; PUTKIRANTA; TOIVANEN, 2012; LAUGEN et al., 2005; LEE et al., 2005; ULUSOY; IKIZ, 2001; VOSS et al., 1995)	1. To identify the practices to be analysed; 2. To build a tool to evaluate the degree of implementation of each practice; 3. To collect data; 4. To analyse data and identify correlations between the practices implementation and companies' performance, so it is possible to identify the best practices.	• Qualitative aspects captured as quantitative data (scale items); • Survey approach; • Adoption of a set of practices as previous reference; • Statistical analysis to stablish correlation between items.
(COOPER; EDGETT; KLEINSCHMIDT, 1999, 2004a, 2004b; FOLZ, 2001; GRIFFIN, 1997; KAHN; BARCZAK; MOSS, 2006; MENKE, 2013; OZER; CHEN, 2006; RAGATZ; HANDFIELD; SCANNELL, 2003)	1. To identify the key business processes that will be the benchmarking focus and the practices related to them; 2. To build a tool to evaluate the degree of implementation of each practice; 3. To collect data; 4. To establish the criteria to set companies into comparison groups; 5. To compare the groups and identify the practices that are adopted by best performers that are not adopted by the comparison group.	• Qualitative data captured as quantitative data (scale items); • Survey approach; • Adoption of a set of practices as previous reference; • Direct comparison between groups of companies.

(DELBRIDGE; LOWE; OLIVER, 1995)	1. To establish the criteria to select companies and set them into comparison groups; 2. To collect data about the companies using a structured interviews and expert panels 3. To analyse the data by systematically comparing the groups; 4. To identify the practices that are adopted by best performers that are not adopted by the comparison group.	• Qualitative data focus; • Direct comparison between groups of companies.

(AMADO; SANTOS; SEQUEIRA, 2013; DAI; KUOSMANEN, 2014; DESPOTIS, 2005; HORTA; CAMANHO; MOREIRA DA COSTA, 2012; MORAIS; CAMANHO, 2011; RUIZ; SEGURA; SIRVENT, 2015; SHERMAN; LADINO, 1995)	1. To establish the sample and scope of study; 2. To propose a method to identify the relative performance/ efficiency of the units in the sample using a mathematical model; 3. To define the parameters and variables required by the model considering data availability and objective of the study; 4. To rank or identify the best performance units among the sample; 5. To excerpt from the model the parameters and variables that impact the most on the best performance, usually applying regression analysis.	• Quantitative data focus; • Mathematical programming based approach.

Source: Author (2016).

Similar methods were aggregated accordingly to its general methodological approaches, so one may find slightly different steps in a particular study when compared to the steps presented in the Table 9. Although some engage a wide variety of problems with some methodological variations, it was considered that the studies in the same group have adopted (explicitly or not) the same general steps. The discussion is centred in the proposed benchmarking method, not in the research method – some studies proposed a method for benchmarking and have applied a different one to the research itself, for example Szulanski (1996). Some studies used simultaneously different methodological approaches, in those cases, the study was classified in main method focused. The steps regarding preparation and allocation of resources were

ignored to keep the discussion around the main steps of data collection and analysis.

From all the articles identified as relevant at the bibliometric analysis phase of this study, four of them seemed to be less related to benchmarking best practices: Cohen and Levinthal (1990), Barney (1991), Eisenhardt (1989) and Banker et al. (1984). Yet, they were consistently cited by the articles in the database, which indicates that they are relevant seminal articles that influenced this area of research by their conceptual or methodological contributions.

The different methods identified in Table 9 bears similarity to the categories proposed by Davies and Kochhar (2002), who identified three main types of studies that linked practices to performance:

- the *ideal model method*: it measures companies against a list of best practices and performance that has been stablish previously – this are studies such as Kuula et al. (2012), S. H. Lee et al. (2005) and Laugen et al. (2005).
- the *benchmarking method*: the best practices are identified by investigating the companies that are achieving high levels of performance – this are studies such as Camp (1989) and Delbridge et al. (1995).
- the *testing of hypothesis method*: it is focused on testing hypothesis and analysing relationship between practice and performance measures – for examples see Suarez et al. (2016), Wang (2013) and Kuula et al. (2012).

The *ideal model method* is useful to compare different companies and heterogeneous groups, since it is stablished a standard to compare against, but it does not identify unique practices that a company may have that is affecting its performance. In this way, an important issue with this method is that it must regularly review the best practices standards to ensure its validity (DAVIES; KOCHHAR, 2002).

The main advantage of the *benchmarking method* is that it may identify new good practices that may emerge in best performing companies, yet they can't generalize the findings due the lack of data to stablish statistical analysis.

The *testing of hypothesis method* is very important to generalize findings, although they are restricted in focus of analysis and it requires a larger sampling.

Wellstein and Kieser (2011) also proposed a classification for approaches to identify best practices that are quite different from the methods identified in this study:

- Through *expert judgement based on empirical evidence* – common in medicine, this approach is based on a group of experts that screen and evaluate available empirical evidence and stablish consensus on good or best practices. In these approach, it is important that practices are narrowly defined and described in great detail.
- Through *success factor research* – this is one of the most common approaches on management science, yet, the have been questions about its capacity to correctly isolate best practices that are correlated with organizations' performance. Ragatz et al. (2003) is an example of this kind of study.
- Through *benchmarking* – in this approach, the reasoning is to find the practices of the best performers that explain its differences for the second best performer.

Among the methods generally presented in this session, Eisenhardt (1989) and Korpela and Tuominen (1996) stablish ground for case studying in benchmarking, which is the method that one usually adopts to propose first suggestions and hypothesis for further investigation. Thus, perhaps, this kind of method is the one that will first identify potential new *best practices* that sought to be confirmed as *best* in other statistically sound studies.

3.2.3. Notes and remarks

Not all the articles selected to in-depth analysis contributed with distinguished discussions and critiques on the methods usually adopted in the field. The authors presented in this session were considered to be the ones that brought significantly new notes and remarks to the scientific discussion.

Delbridge, Lowe and Oliver (1995) emerges as the first relevant study that addresses important issues and common drawbacks of benchmarking studies. They point that the benchmarking process represents a significant challenge because of the difficulty of making sufficiently precise comparisons.

Benchmarking demands a systematic rigorous approach to data collection with an emphasis on quantitative "hard" information (DELBRIDGE; LOWE; OLIVER, 1995; PHILLIPS; APPIAH-ADU, 1998). The issue of comparability is of central importance to benchmarking studies, as they stand or fall by the legitimacy of the comparisons they make (DELBRIDGE; LOWE; OLIVER, 1995). Wellstein and Kieser (2011) go further on the critique on benchmarking: "brokering practices is easier when there is a story to tell that a superior practice has been found systematically. Benchmarking provides such a story" (p. 687).

Kuula et al. (2012) suggest that there is a lifecycle for the practices used in companies – many of the practices adopted in the late 1990s are already out of date. Such observations make even difficult to run longitudinal studies with relevant results, since the aging of the practices may impact significantly in the analysis. This suggestion was also proposed by Laugen et al. (2005).

Davies and Kochhar (2002) critique several studies that related best practices and performance, calling attention to the following points:

- The most of the studies remain descriptive, perhaps because the difficulty to attach mathematical relationships to companies' environments, which have many variables impacting in;
- A large proportion of the studies relates the effects of best practices on performance;
- There are studies demonstrating that some practices may be prerequisite for others more sophisticated, which suggests that studies of best practices should not only consider the extend of relationship with performance, but also with other practices;
- It is important to consider the national and sector context when analysing the adoption of a best practice.

Laugen et al. (2005) addressed some of this issues in their study, using ANOVA and regression analysis, they identified the practices adopted by best performers that was related to their superior results. Similarly, S. H. Lee et al. (2005) also identified the best practices in construction industries.

Wellstein and Kieser (2011) presented more critiques to researches on "best practices". They argue that, initially, early research is criticized

for its simplistic methodology, then, the most sophisticated studies come up with results that contradict earlier analysis, and because of that, lacking convergence of studies, it gets difficult to establish a relationship between the contingencies identified in diverse academic studies and the specific conditions that prevail in a company looking for a superior practice. Moreover, these studies can only analyse historical data, assuming that practices that worked in the past will continue to do so in the future (WELLSTEIN; KIESER, 2011), which is an assumption that has already been contested by Kuula et al. (2012).

Critiques such as those may have lead the researchers to look for more quantitative approaches, which could explain the increase adoption of DEA applications in benchmarking studies in the last ten years – although not all DEA studies ground their research in managerial benchmarking cluster authors, there are some that do, such as Forker & Mendez (2001), which suggests that hypothesis.

DEA may help researchers and practitioners to identify the benchmark units in a given set, but there are practical and methodological aspects that must be considered. One of the most relevant issues pointed by Cook et al. (2014) refers to the little attention that usually is paid to insuring that the selected measures properly reflect the process under study – and even when it is the case, one can never be completely assured that all of the relevant variables have been considered. Another point of attention on DEA method is the presence of outliers on the dataset, which could severely affect the DEA frontier, that is very sensible to extreme observations (MORAIS; CAMANHO, 2011).

Despotis (2005) argues that during the optimization process DEA selects the weights to aggregate the indicators in favour of the DMU in such a way that the unit scores the best efficiency possible. Thereby, a low DEA score undoubtedly show a poor performance unit regardless of the weighting scheme selected to aggregate inputs and outputs. Nevertheless, Morais & Camanho (2011) argue that it also can be a weakness of DEA models, as it allows some indicators to be assigned a zero weight, which means that some factors can actually ignored in the performance assessment. To overcome this situation one can impose weight restrictions on the DEA model as one can notice in the work of Ruiz et al. (2015).

In order to define a common set of outputs and inputs, DEA assumes that the DMUs are undertaking similar activities, producing comparable products, have a similar range of resources available, and

operate in similar environments (RUIZ; SEGURA; SIRVENT, 2015). This issue has not been addressed by the most of the DEA studies analysed in this study.

Dai and Kuosmanen (2014) recognize that there may be relevant differences between units identified as benchmarks from the evaluated DMU (the differences can refer to the input profile, the output structure and the scale sizes), and also that DEA is sensitive to random noise, heterogeneity of units, and differences in their operating environment. Thus, "while DEA can identify successful units, it may be difficult to transfer the success recipes to inefficient DMUs if the success is due to external conditions or just good fortune" (DAI; KUOSMANEN, 2014, p. 180).

In contrast, there are some works that put some light on that issue. The first relevant study that addressed the linking between best practice and performance used DEA to analyse the efficiency on implementation of total quality management practices by comparing the results obtained by DMUs with the reported effort to implement the practices – see Forker & Mendez (2001). More recently, Ruiz et al. (2015) offer an alternative DEA method that allows the identification of close benchmarks for each DMU, which gives specific context-related best practices for the DMU, while Amado et al. (2013) apply DEA alongside with Mann-Whitney rank statistics and Malmquist Productivity Index to investigate whether differences in DMUs' efficiency can be attributed to a particular managerial programme or to innovation. The identification of the impact of a best practice occurs when a group of DMUs that have implemented it show an increase in productivity between two different periods, while the identification of innovations is done by analysing the DMUs that have contributed to shift the frontier to higher levels (AMADO; SANTOS; SEQUEIRA, 2013).

3.3. CONCLUDING REMARKS

This session aimed to overview the scientific literature on benchmarking of best practices, tracking the most relevant articles, authors, journals, and understanding the main concepts, methods and discussions on the field.

The adopted methodological procedures consisted in procuring a dataset from the *Web of Science* and running bibliometric analysis using two computer program that allowed identify the main articles through

three citations criterion. Using these different combined criteria proved to be useful in terms of avoiding bias. By taking into consideration the clusters of the co-citation analysis it became evident the existence of two different groups of authors and publications: one related specifically to DEA, and other that presented a wide variation of methods and approaches. Moreover, the inclusion of novel articles in the pool for further in-depth analysis, allowed this study to cover the most updated discussions on the benchmarking of best practices.

The overall resulting relevant articles identified in bibliometric analysis seemed to be well balanced in terms of novelty and relevance. The citation analysis had a varied outcome, with articles ranging from 1989 to 2014 and with more than 17% of them from earlier than 2010. In the other hand, the relevant papers identified in the two main clusters of the co-citation map tended to be older, with about 90% of them older than 2000. Both analysis has shown to be complementary, about 50% of the relevant articles identified in the co-citation map were also identified in the citation analysis.

The analysis of scientific output substantiates the increasing in production of articles in this research topic. 47% of the articles in the dataset corresponds to the production of the last five years (between 2010 and 2015), and the year of 2015 alone represents more than 10% of all papers in the dataset.

Despite this increasing amount of studies been published related to benchmarking of best practices, there are relevant critique that must be acknowledge. The critique put in doubt the methods adopted and go further questioning the very existence of practices that can be recognized as “best”.

To counter the critiques on benchmarking studies, there are new DEA studies that propose novel methodologies that may address with statistical soundness the relation of best practices and companies’ superior performance.

4. ANALYSIS OF A BEST PRACTICES ASSESSMENT INSTRUMENT USING ITEM RESPONSE THEORY

Although there are several ways to analyse and validate questionnaires in operations research, it is not common to see papers concerned with advocating the validation of their instruments and proposing lessons learned with the quality of their questionnaire.

What one can usually find is pilot tests of questionnaires, or validation by specialists. The problem with these approaches is that they give little empirical information on how an item is better than other, or even whether the proposed construct has been correctly inquired. More information on questionnaire validation can be found in Straub (1989), Mackenzie *et al.* (2011) or Kuula and Putkiranta (2012).

From that perspective, this session aims to present how the Item Response Theory (IRT) was applied to analyse and validate a questionnaire for best practice assessment, also providing some insights on the lessons learned in this process.

The idea to apply the IRT to validate the questionnaire occurred during development of the questionnaire when it became necessary to understand how difficult each item was and how good they were at differentiating the assessed companies. Understanding the quality of the items permitted to get better insights on what makes a good item and how to balance the difficulty of the questionnaire.

The IRT has been applied to measure latent traits of individuals which cannot be observed directly (such as intelligence, knowledge of a certain subject, or companies management maturity) by defining the probability of an individual to give specific answers to the items of a questionnaire (ANDRADE; TAVARES; VALLE, 2000a; BALBIM JUNIOR; BORNIA, 2011; RASCH, 1993; STOCKING; LORD, 1983).

One of the objectives of the IRT is to estimate the characteristics of the items of a questionnaire using probabilistic models based on the responses of a population sample. The items are characterized by three parameters: 1. their capacity to discriminate respondents with different proficiencies; 2. their difficulty; and 3. the probability of randomly answer it correctly. Based on these item characteristics, the IRT will score each respondent.

The IRT has already been applied in the development of questionnaires related to operations and management research fields, as

shown by Trierweiller *et al*. (2012), Vasconcelos *et al.* (2013), Alexandre *et al.* (2002) and Vargas *et al.* (2008) who have applied IRT to validated questionnaires that measure management practices adopted by organisations in different contexts and fields.

Alexandre *et al.* analysed the adoption of total quality management practices by industrial companies using the three-parameter logistic model considering dichotomized responses (whether the practices were implemented or not), and, in doing so, they have analysed the characteristics and the quality of the items of their questionnaire.

Vargas *et al.* proposed a questionnaire and a scale to measure the practices of management of intangibles in organizations so they could compare both intangibles and organizations performance and identify the aspects that give competitive advantage to them.

Trierweiller *et al*. verified the validity of using IRT to measure evidences of adoption of environmental management practices among industrials. This approach was particularly useful to evaluate constructs that were difficult to observe otherwise, concluded the authors.

Vasconcelos et al. also applied IRT in the organizational context to propose and validate a scale to measure the success of micro and small business in a regional context, backing Trierweiller et al. and Alexandre's et al. conclusion that this method could be applied in validation of tests and questionnaires.

None of these studies emphasise the validation processes nor describe in more depth the reasoning behind its analysis. Considering that, the main contribution of this study is to propose procedures and rational to analyse the items of the instrument and identify aspects that can be improved in the design of questionnaires.

4.1. THE QUESTIONNAIRE FOR BEST PRACTICES ASSESSMENT

The questionnaire for assessment of best practices of industrial plants (IEL/SC, 2010) used in this study was developed in 2010 by Instituto Euvaldo Lodi (IEL/SC) [3].

The questionnaire is used to assess the maturity of companies in adoption of management and operations practices and it allows companies to compare the adoption of the practices among others in the same industry or in the same region, so they can identify improvement opportunities and enhance its productivity and competitiveness.

The questionnaire covered subjects related to production management, quality, health and safety, environment, costs and finance, strategy planning, innovation, social responsibility and marketing. It consists of 46 items, of which 24 assess practices adoption (example in Figure 13) and 22 assess the company's performance (example in Figure 14), such as client satisfaction, failure rates, profits and so on.

As shown in the examples, each item is composed of an identification code (such as PP3 or PP11), a name, a general description below the name, a scale (from 1 to 5) and descriptions for scores one, three and five.

For the items that assess practice adoption, in the descriptions for scores three and five there are two or more requirements that the company must fulfil in order to get that score. If the company has some of the requirements of a score, but not all of them, it will score two or four. If the company has none of the requirements, it will score one – so the description of this score is always indicating that the company did not implemented that practice at all.

As shown in the examples, each item is composed of an identification code (such as PP3 or PP11), a name, a general description below the name, a scale (from 1 to 5) and descriptions for scores one, three and five.

[3] The development of the questionnaire is described at Castro (2010). IEL/SC is a non-profit organization that is part of the Industrial Federation of Santa Catarina State, Brazil.

Figure 13 – Example of an item to measure practice adoption by companies.

PP 3	**Processes mapping and control**	**1**	**2**	**3**	**4**	**5**
How the company ensures that its processes are adequate to meet the requirements of the products? Are the processes mapped? Are there work instructions? Are there inspections checkpoints? What are the devices used to measure the quality of products? Is it possible to track products from their raw materials, troughout their processes and destination?		The processes are not mapped and there are no established control mechanisms.		The company has mapped and defined work instructions for production processes, identifying inputs, steps, and outputs. The processes meet the requirements identified for the product. The control occurs through the work instructions and inspection checkpoints. There is devices and/ or standards to measure the quality of products.	The sales, delivery and key support processes are also mapped. There are indicators for monitoring and control the processes. Audits are carried out to check compliance of processes. There are mechanisms that ensure the identification and traceability of the products. There are improvements implemented through critical analysis of processes' performance. All measuring devices are calibrated periodically and certified by third parties.	

Source: IEL/SC (2010), translated by author (2016).

Figure 14 - Example of an item to measure performance of companies.

PP 11 **Customer satisfaction**	**1**	**2**	**3**	**4**	**5**
Average customer satisfaction ratings, measured last year.	The satisfaction level is below 60%. This indicator is not measured.		Customer satisfaction level is between 70% and 80%.		Customer satisfaction is over 90%.

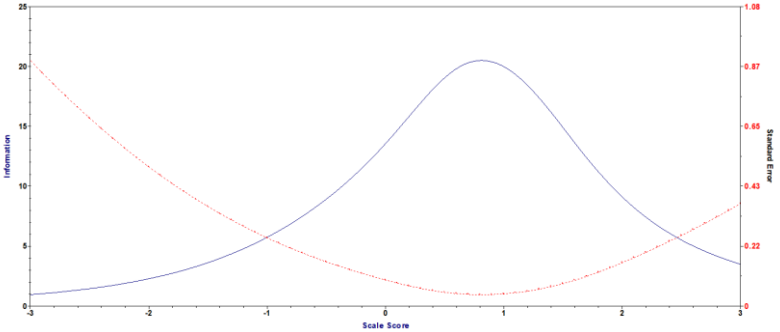
Source: IEL/SC (2010), translated by author (2016).

For the items that assess practice adoption, in the descriptions for scores three and five there are two or more requirements that the company must fulfil in order to get that score. If the company has some of the requirements of a score, but not all of them, it will score two or four. If the company has none of the requirements, it will score one – so the description of this score is always indicating that the company did not implemented that practice at all.

4.2. ITEM ANALYSES

The total test information curve, presented in Figure 15, shows that in general the questionnaire is better at scoring companies with maturity ranging approximately between -0.5 and 2.0. This may also show that this questionnaire is difficult for the companies in the sample.

Figure 15 – Questionnaire total information curve.



Source: Author (2016), plotted in *Bilog-MG*.

Analysing the difficulty per subject of the questionnaire, presented in Figure 16, it becomes evident the differences on the assessment of management practices and company's performances. *People management* and *Health and safety* are two of the subjects that are much easier to get a good score in performance than in practice, while *Environment* and *Innovation* are the opposite, it is easier to score better in performance than in practice.

Figure 16 – Difficulty scores per questionnaire subject.

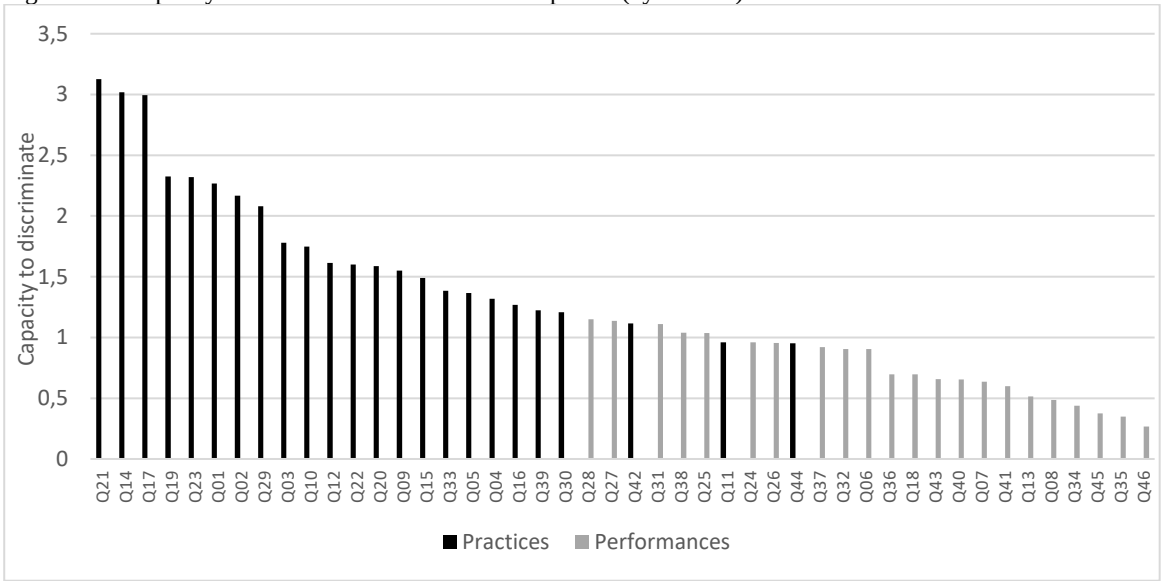
Source: Author (2016).

The Figure 17 presents the parameter *a*[4] of the items, which means their capacity to discriminate companies with different management maturity.

In general, the items that measure performance are not efficient in differentiating the companies. The scales used in the performance indicators must be reviewed. Another aspect of the item construction that may be impacting on this difficulty to differentiate companies is a requirement that demands that the company measure some indicators (see Figure 14, for example), so, companies that does not measure them cannot be differentiated.

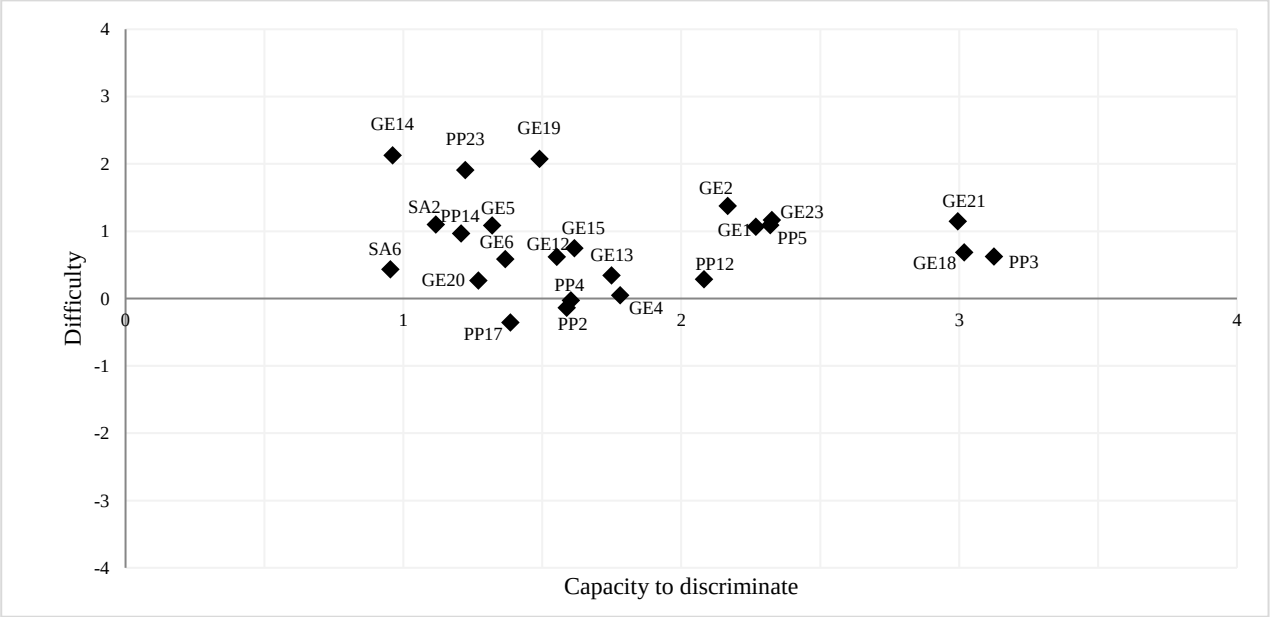
[4] Parameter *a*: Capacity of the item to discriminate respondents with different latent traits. See Table 3 in page 33.

Figure 17 – Capacity of the items to discriminate companies (by authors)



Source: Author (2016).

Figure 18 – Items distributed by their difficulty and capacity to discriminate.



Source: Author (2016).

The Figure 18 presents the items distributed by their capacity to discriminate companies and their difficulty. This graph helps identifying the characteristics that make a good item, for instance, when selecting two items with similar difficulty, such as SA2 and GE21, one must try to understand why one item is much better in discriminating the companies than the other.

It is important to notice the differences between the difficulty and the capacity to discriminate of the items. The aspects that make an item easier or harder are not necessarily the same that make an item good or bad to discriminate companies with different management maturity, as shown in the Figure 18.

The Table 10 presents the ten items with lowest capacity to discriminate the companies and the analysis made by the experts panel that discussed the items' quality, noticing that although this are the "worst" items, yet they are fairly good at differentiating the companies – they have a discrimination parameter close or above 1,0.

To illustrate the expert's analysis, the items SA2 (Figure 19) and GE21 (Figure 20) are presented next, so one can notice the differences between them and what makes GE21 much better than SA2 to differentiate a company's management maturity.

The main issues that the experts' panel identified to impact on the item's quality is related mainly to how clear are the item requirements, how easy it is to verify the compliance of the requirements, and how well distributed the requirements are between the scenarios three and five.

In conclusion, the items that measure practice are good and they can evaluate the management maturity of the companies with a good degree of certainty, although the questionnaire in a whole is slightly difficult for the assessed companies. It is recommended that a couple of items could be made easiest. In the other hand, the items that measures performance are not balanced, so the scales adopted must be reviewed.

Table 10 – Experts panel analysis on the ten items with lowest capacity to discriminate companies.

Item	Subject	Capacity to discriminate	Difficulty	Require-ments in scenario 3	Analysis
SA6	Incident investigation	0,95	0,43	2	The item is too easy, most part of the companies have implemented this practice, so it does not differentiate them. The requirements description must be clearer.
GE14	Marketing planning	0,96	2,12	2	The item is too hard, most part of the companies did not have implemented it, so it does not differentiate them.
SA2	Environmental aspects, impacts and controls	1,12	1,10	4	There are too many requirements on the scenario 3, covering a wide range of management practices. This item should be separated in two.
PP14	Equipment maintenance	1,21	0,97	3	The requirements descriptions must be clearer.
PP23	Ideias capturing	1,22	1,91	2	One of the requirements of the scenario 3 is too demanding. Description must be clearer.
GE20	Salary and benefits	1,27	0,26	2	The item is too easy, most part of the companies have implemented this

					practice, so it does not differentiate them.
GE5	Projected cash flow	1,32	1,08	3	The adoption of this practice is less correlated with the companies' maturity than others.
GE6	Cost management	1,37	0,58	3	There are some difficult concepts related to this item that make its requirements descriptions not so clear.
PP17	Flexibility at work	1,38	-0,36	2	It is difficult to ascertain the compliance of the requirements of this item.
GE19	Personnel performance evaluation	1,49	2,07	5	Too many requirements on the scenario 3. Item is too hard.

Source: Author (2016).

Figure 19 – Example of item that do not discriminate companies well.

SA 2 **Environmental aspects, impacts and controls**	**1**	**2**	**3**	**4**	**5**
Which are the environment impacts of the company? Environmental aspect: element of the activities or products of an organization that can interact with the environment. (ISO 14001) Environmental impact: any change in the enviornment adverse or beneficial, resulting, in whole or in part, from the environmental aspects of the organization. (ISO 14001)	There is no formal procedure for the identification of environmental aspects and impacts and controls ere not in place to handle them.		There are formalized and disseminated procedures for identifying environmental aspects. Aspects that cause significant impact on the environment are identified. The environmental impacts of the production processes are controlled. There are indicators to monitor the performance of established controls.		Aspects related to products and services are identified, in addition to those related to production activities. There are aspects identified from changes in the company (change management). The environmental impacts are controlled covering design, production, use and disposal (product life cycle). There are formalized procedures for identifying and responding to potential environmental emergencies. The employee participates in identifying environmental aspects and impacts, as well as setting its controls.

Source: IEL/SC (2010), translated by author (2016).

Figure 20 - Example of item that discriminate companies well.

GE 21 **Training**	**1**	**2**	**3**	**4**	**5**
Is there any personnel training planed? How the training needs are identified?	There is no formal training plan.		There is a systematic plan for training. There are records of regular monitoring of the training plan. The necessary skills are mapped to meet the product quality. There are practices to train and integrate new employees.		The needs are mapped to meet customer needs, organizational strategies and people's needs, in addition to product quality. Trainings are evaluated for their effectiveness. There is efforts to develop both technical and behavioral competences.

Source: IEL/SC (2010), translated by author (2016).

4.3. CONCLUDING REMARKS

In this session, the item response theory was used to analyse 302 responses for a questionnaire with 46 items that assess productivity and competitiveness of small and midsized industrial companies, identifying the aspects that affected the quality of the questionnaire.

By the errors presented in the calibration of the items, it was possible to assume that the parameters were successfully estimated with a good margin of certainty.

Considering the items and questionnaire analysis results. The items that measure practice adoption were found to be good at assessing management maturity of the companies with a good degree of certainty, although the questionnaire in a whole is slightly difficult for the assessed companies. In the other hand, the items that measure performance are not efficient in differentiating the companies and the scales references used in the performance indicators must be reviewed

The main issues that the experts' panel identified to impact on the item's quality are related mainly to:

- how fit is the scale of the item to the average company management maturity;
- how clear are the item description;
- how easy it is to verify the compliance of the requirements of the items by the companies; and
- how well distributed are the requirements between the scenarios three and five.

The main focus of this study was to emphasise the process to validate que questionnaire and describe in more depth the reasoning behind the analysis, proposing novel procedures and rationale to analyse the items of the instrument and identify aspects that can be improved in the design of questionnaires.

The IRT has shown to be a useful instrument that can help researchers to analyse and understand the quality of their questionnaires and learn with the respondents how good their items are. This kind of knowledge can be vital to one that is willing to design better research instruments.

5. A FRAMEWORK TO IDENTIFY AND MEASURE BEST PRACTICE ADOPTION

Generally, the papers presented in this study address the identification of best practices and benchmarking processes with a wide range of methods and applications.

Schiele and Krummaker (2011), Camp (1989), Drew (1997), Korpela & Tuominen (1996) and Szulanski (1996) focus on the benchmarking process by the point of view of a company, proposing methods that can be applied in order to identify and implement improvements in its own process. Yet, they pay little attention to the identification of the benchmark – the best performer to whom one may want to compare with – assuming that the best performers are well known and their practices are the best practices.

In the other hand, Voss et al. (1994) address the benchmarking process and the identification of best practices not from the point of view of a company, but as someone intended to run the process systematically in several sites. This approach, with the application of their method widely, resulted in the construction of a database of practices and related performances that permitted later studies to identify correlations of practices and performance (HANSON; VOSS, 1995; VOSS; ÅHLSTRÖM; BLACKMON, 1997).

Although Voss and his colleagues were able to build a database of practices in-depth analysis, the questionnaire they used to assess the companies was built from the *Malcom Baldridge Award* practices and literature examination, what means that they already had a pool of practices that they intended to verify. This represents a significant bias on the research, because (1) the performance of the companies could be explained by other practices that were not assessed, and (2) the adoption of the assessed practices by every company regardless adaptation could be considered a drawback as Voss himself discussed in a later paper (SOUSA; VOSS, 2008).

Specifically about which practices should be assessed, Delbridge, Lowe and Oliver (1995) and Collins (2001) works provide an interesting approach on how to investigate the practices that the best performers have that differentiate them from the rest. In the other hand, these works do not consider the adoption of the identified differentiation elements by a broader sample of companies in a way that it would be possible to

demonstrate statistically the relation of those elements to a superior performance.

Wellstein and Kieser (2011) argue that the research on benchmarking of best practices has not been able to deliver results based on uncontested methods.

Initially, early research was criticized for its simplistic methodology, then, the most sophisticated studies come up with results that contradict earlier analysis – see Forker and Mendez (2001), Laugen et al. (2005), Lee et al. (2005) and Ulusoy and Ikiz (2001).

Because of lack of convergence, it became difficult to establish a relationship between the contingencies identified in diverse academic studies and the specific conditions that prevail in a company looking for a superior practice (WELLSTEIN; KIESER, 2011).

Moreover, these studies could only analyse historical data, assuming that practices that worked in the past will continue to do so in the future (WELLSTEIN; KIESER, 2011) – an assumption that has already been contested by Kuula et al. (2012).

Critiques such as those may have lead the researchers to look for more quantitative approaches, which could explain the increase adoption of DEA applications in benchmarking studies in the last ten years - see Sherman and Ladino (1995), Morais and Camanho (2011), Dai and Kuosmanen (2014), Horta et al. (2012), Despotis (2005), Ruiz et al. (2015) and Amado et al. (2013).

DEA may help researchers and practitioners to identify the benchmark units in a given set, but there are practical and methodological aspects that must be considered. One of the most relevant issues pointed by Cook et al. (2014) refers to the little attention that usually is paid to insuring that the selected measures properly reflect the process under study – and even when it is the case, one can never be completely assured that all of the relevant variables have been considered. Another point of attention on DEA method is the presence of outliers on the dataset, which could severely affect the DEA frontier, that is very sensible to extreme observations (MORAIS; CAMANHO, 2011).

Dai and Kuosmanen (2014) recognize that there may be relevant differences between units identified as benchmarks from the evaluated DMU (the differences can refer to the input profile, the output structure and the scale sizes), and also that DEA is sensitive to random noise, heterogeneity of units, and differences in their operating environment.

Thus, “while DEA can identify successful units, it may be difficult to transfer the success recipes to inefficient DMUs if the success is due to external conditions or just good fortune” (DAI; KUOSMANEN, 2014, p. 180).

Moreover, an important drawback of these DEA studies is that they are always limited by the data available – to a best practice be identified it must have been assessed and data must be collected previously – one can never guarantee that the data available covers all the possible best practices there is. In order to do that, it would be necessary a flexible method, with a flexible questionnaire that can evolve by assessing new practices and discarding obsolete ones at the same time that it permits the comparison between units.

Considering that, the Item Response Theory (IRT) may be an alternative for these drawback because it permits flexibility to change the questionnaire with new practices without losing its capability to compare units. One can remove and add new items without changing the final score of the respondents, allowing the comparison between them, even if they responded different questions.

Therefore, the objective of this study is to present a method to identify best practices among a set of companies that is adaptable and statically relevant.

5.1. THE BEST PRACTICE IDENTIFICATION FRAMEWORK PROPOSITION AND TEST

The framework proposed in this study is presented below, followed by the questionnaire analysis results.

5.1.1. The proposed framework

The discussion presented in the literature review session suggests that there is not one reference method to identify best practices and conduct a benchmarking process, instead, a combination of them may help avoiding the drawbacks of each one particularly, resulting in a more complete method to be used in further researches.

In this manner, the framework proposed by this study had to fulfil some premises that were founded on the lessons learned from the literature:

- The assessment of the elements that differentiate the best performers from the rest must take into account qualitative aspects;
- The qualitative aspects must be measured in quantitative metrics in a way that it becomes possible to tabulate and analyse them statistically;
- It must be possible to collect enough data to evince correlations between the measured variables;
- It must permit flexibility in adding or excluding practices to be evaluated on the process without compromising the assessment scale and the possibility to compare companies.

The Table 11 presents the *Framework to Identify Best Practices*.

Table 11 – The proposed *Framework to Identify Best Practices*.

Step	Context	Reference methods
1. To define the scope of analysis and processes where to find the best practices	The definition of scope may consider a specific area of interest or may be generalist covering the whole organization.	(VOSS; CHIESA; COUGHLAN, 1994)
2. To identify the good practices associated to the processes	The identification of good practices may occur in with any method – case research, benchmarking, survey, literature review.	(COLLINS, 2001; EISENHARDT, 1989; GUPTA; VERMA; VICTORINO, 2006; KORPELA; TUOMINEN, 1996; MACKENZIE; PODSAKOFF; PODSAKOFF, 2011; VOSS; TSIKRIKTSIS; FROHLICH, 2002)

3. To develop a questionnaire with items that evaluate the extent to which each good practice is implemented, and the performance that outcomes from the adoption of these practices	In this step, it is important that the constructs identified in the step two are translated into scale items. The lessons learned in the session 0 may be useful in this moment.	(MENKE, 2013; VOSS; CHIESA; COUGHLAN, 1994)
4. To collect data	It is strongly recommended that the data collection is made with the researcher visiting and assisting the company understanding the items.	(VOSS; CHIESA; COUGHLAN, 1994)
5. To feedback the assessed company	The company will be looking for a feedback on the status of its practice compared with the available database.	
6. To check the rise of new practices during the data collection	If a company has developed a new practice, it must be registered for further analysis.	(COUGHLAN; COGHLAN, 2002; EISENHARDT, 1989; VOSS; TSIKRIKTSIS; FROHLICH, 2002)
7. To analyse data and verify whether each good practice assessed is related to a superior performance	This step may follow a DEA approach or other statistically sound analysis.	(AMADO; SANTOS; SEQUEIRA, 2013; KUULA; PUTKIRANTA, 2012; LAUGEN et al., 2005; RUIZ; SEGURA; SIRVENT, 2015)
8. To calibrate the items and adjust the questionnaire	In this moment the researchers can adjust or exclude items accordingly to its relevance and quality, based on the IRT.	(ALEXANDRE et al., 2002b; ANDRADE; TAVARES; VALLE, 2000b; PACHECO; ANDRADE; BORNIA, 2015)

9. To review the necessity to add new items for novel practices identified	The information collected in the step 6 is now analysed by the researchers that may propose a new practice and build new items to assess it in a new application of the questionnaire.

Source: Author (2016).

The method was considered feasible, once it was proposed based on a data collection process that is already in place, executed by IEL/SC. In order to implement this proposition, it would be necessary to adjust the current process. Moreover, other consulting companies could adopt the proposed steps in order to improve its own instruments for diagnose of best practices.

With this method, there is an attempted proposition to address all the critique of assessment of best practices identified in the literature:

Critique: there is a lack of correlation between practice and performance. The collection of data will permit to stablish correlations between practice adoption and performance, as proposed by Voss, Åhlström and Blackmon (1997). It will also be possible to test the contingencies and the context in which each practice drives performance, as proposed by Amado, Santos and Sequeira (2013) and Ruiz, Segura and Sirvent (2015).

Critique: the studies that analyse correlation of practices and performance usually assumes a set of best practices as reference. The method starts with a comprehensive analysis that accepts a wide variety of methods: case studying (EISENHARDT, 1989; VOSS; TSIKRIKTSIS; FROHLICH, 2002); success factors analysis (KORPELA; TUOMINEN, 1996); empirical research (GUPTA; VERMA; VICTORINO, 2006); construct measurement (MACKENZIE; PODSAKOFF; PODSAKOFF, 2011); and control group analysis (COLLINS, 2001).

Critique: the studies that correlate practices and performance, alongside with the DEA application studies, are dependent of data availability. The proposed method permit address a wide range of practices – the instrument tested had 46 items. The only limitation to

comprehensiveness of analysis is the research scope. Additionally, the steps 1, 8 and 9 of the method address the aspects and practices that may have been neglected.

Critique: the best practices may expire and become outdated. The steps 7 and 9 can deal with the analysis of the practices that are driving performance and if there are new practices that should be considered in the analysis. These analyses can be done as proposed by Amado, Santos and Sequeira (2013), Kuula and Putkiranta (2012), Laugen et al. (2005) or Ruiz, Segura and Sirvent (2015).

5.1.2. Questionnaire analysis

The Table 12 presents the results of the IRT analysis for all the 46 items of the questionnaire. Alongside with the slope and threshold parameters, in order to compare the correlation of each item to respondent final score with the IRT's parameters, it is presented the classical analysis of *Person* and *Biserial* coefficients.

The results show that in general the items are well calibrated and can assess the management maturity of companies with a good degree of certainty.

The items that measure practices demonstrated to have a good correlation with companies' management maturity. *Mapping and process control, Recruiting and selection, Infrastructure and workplace, Training, Strategic basis* are the items that presented the higher correlation to a company maturity: *Pearson* above 0.6, *Bisereal* above 0.8.

Considering the capacity to differentiate companies with different management maturity, the best items (Slope above 2.0) are *Mapping and process control, Recruiting and selection, Training, Structure of positions and functions, Problem analysis and improvement actions, Strategic basis, Strategic planning and Infrastructure and workplace*, since companies that scores higher on those items will probably have a higher management maturity.

There are opportunities to improve some items, mainly those that measure performance, since they are not so efficient in differentiating the companies. The scales used in the performance indicators must be reviewed. Another aspect of the item construction that may be impacting on this issue is a requirement that demands that the company measure

some indicators, so, companies that does not measure them cannot be differentiated.

Table 12 – Items analysis results.

Item	Subject	Type	Pearson	Biserial	Slope (*a*)	S,E,	Thres-hold (*b*)	S,E,
Q01	Strategic basis	Practice	0,596	0,849	2,268	0,352	1,064	0,103
Q02	Strategic planning	Practice	0,546	0,842	2,166	0,355	1,374	0,136
Q03	Financial control	Practice	0,562	0,704	1,780	0,236	0,045	0,083
Q04	Projected cash flow	Practice	0,445	0,605	1,318	0,184	1,082	0,154
Q05	Cost management	Practice	0,478	0,614	1,366	0,199	0,583	0,108
Q06	Costs reduction	Performance	0,359	0,476	0,904	0,157	1,184	0,203
Q07	Profitability	Performance	0,274	0,343	0,635	0,126	0,156	0,199
Q08	Return on investment	Performance	0,183	0,230	0,486	0,110	0,396	0,267
Q09	Market knowledge	Practice	0,540	0,699	1,551	0,206	0,618	0,109
Q10	Communication with clients	Practice	0,528	0,669	1,749	0,239	0,340	0,089
Q11	Marketing planning	Practice	0,326	0,500	0,961	0,182	2,123	0,343
Q12	Sales structure	Practice	0,518	0,682	1,615	0,205	0,746	0,105
Q13	Sales variation	Performance	0,209	0,262	0,516	0,107	-0,165	0,234
Q14	Recruiting and selection	Practice	0,630	0,841	3,017	0,542	0,683	0,069
Q15	Personnel performance evaluation	Practice	0,357	0,623	1,489	0,274	2,073	0,261

Q16	Salary and benefits	Practice	0,466	0,587	1,269	0,176	0,263	0,111
Q17	Training	Practice	0,606	0,905	2,995	0,479	1,144	0,092
Q18	Turnover of employees	Performance	0,304	0,383	0,697	0,121	0,415	0,189
Q19	Structure of positions and functions	Practice	0,543	0,796	2,326	0,339	1,165	0,103
Q20	Clients requirements	Practice	0,513	0,644	1,587	0,198	-0,138	0,092
Q21	Mapping and process control	Practice	0,677	0,894	3,125	0,522	0,621	0,066
Q22	Acquisition and relationship with suppliers	Practice	0,517	0,648	1,602	0,233	-0,029	0,090
Q23	Problem analysis and improvement actions	Practice	0,579	0,831	2,320	0,322	1,086	0,108
Q24	Quality in the delivery of suppliers	Performance	0,388	0,489	0,959	0,159	-0,365	0,145
Q25	Customer complaints rate	Performance	0,414	0,521	1,035	0,161	-0,206	0,132
Q26	Internal defects	Performance	0,383	0,527	0,954	0,181	1,440	0,243
Q27	Defects costs	Performance	0,448	0,594	1,137	0,189	0,989	0,162
Q28	Customer satisfaction	Performance	0,419	0,539	1,149	0,187	-0,664	0,142
Q29	Infrastructure and workplace	Practice	0,624	0,790	2,081	0,263	0,286	0,078
Q30	Equipment maintenance	Practice	0,471	0,625	1,207	0,198	0,966	0,155
Q31	Material processing time	Performance	0,408	0,669	1,111	0,459	2,313	0,907

Q32	Production lead time	Performance	0,353	0,477	0,906	0,155	1,349	0,240
Q33	Flexibility at work	Practice	0,477	0,603	1,385	0,205	-0,359	0,107
Q34	Storing	Performance	0,163	0,209	0,438	0,097	1,234	0,384
Q35	Inventory turns	Performance	0,089	0,112	0,349	0,086	-0,946	0,411
Q36	On-time deliveries	Performance	0,291	0,366	0,697	0,130	-0,292	0,184
Q37	Time of preparation of equipment or line change	Performance	0,384	0,483	0,920	0,149	0,257	0,150
Q38	Productivity	Performance	0,400	0,501	1,039	0,157	0,120	0,126
Q39	Ideas capitation	Practice	0,363	0,570	1,222	0,222	1,908	0,263
Q40	Ideas generation	Performance	0,188	0,284	0,654	0,147	2,760	0,589
Q41	Return of new products	Performance	0,220	0,297	0,599	0,127	1,850	0,413
Q42	Environmental aspects, impacts and controls	Practice	0,423	0,566	1,116	0,189	1,098	0,170
Q43	Waste reduction	Performance	0,240	0,338	0,657	0,131	2,157	0,442
Q44	Incident investigation	Practice	0,403	0,510	0,953	0,148	0,434	0,150
Q45	Absences for illness	Performance	0,120	0,157	0,375	0,096	-2,133	0,612
Q46	Injuries suffered by employees	Performance	-0,017	-0,023	0,267	0,071	-3,012	0,932

Source: Author (2016).

5.2. CONCLUDING REMARKS

The *Best Practices Identification Framework* proposed emerged from the analysis of the studies identified in the literature review.

The application of the IRT permitted a deeper analysis of the data, indicating how the items were behaving with responses and which one presented issues for further investigation. The reasoning to understand how each item should be analysed was also an important contribution to the processes of validation of the questionnaire, translating the outputs of the IRT analysis into clear aspects of the items to be discussed in the experts' panel.

Regarding the questionnaire and the items analysed, the main issues identified that the impact on the item's quality is related to how clear are the item requirements, how easy it is to verify the compliance of the requirements, and how well distributed the requirements are between the items scale.

The items that measure practice can evaluate the management maturity of the companies with a good degree of certainty, although the questionnaire in a whole is slightly difficult for the assessed companies. It is recommended that a couple of items could be made easier. In the other hand, the items that measure performance are not well balanced, so the scales adopted should be reviewed, or, an alternatively these items can be removed from the questionnaire and assessed separately, since they don't measure the management maturity *per se*, so they can be related to a different latent trait.

The framework test demonstrated its applicability and opportunities, contributing to the discussion of the usage of statistical tools, especially the IRT, to the assessment of best practices and benchmarking purposes.

Perhaps the main opportunity that the IRT presents is the possibility to gradually change the questionnaire in the next applications without losing the data and analysis made. For instance, for future applications, the questionnaire can have different items to measure performance, as long as a part of the new questionnaire is built with calibrated items so a company assessed can have its latent trait (management maturity) measured and benchmarked while data for new items is collected and processed so they can also be calibrated.

In this way, IRT permits not only to enhance items, but also to include items into the benchmark instrument to access new practices that may emerge in the future, maintaining the test scale and ultimately allowing the comparison of companies that have answered different questionnaires.

6. CONCLUSION AND FUTURE RESEARCH

This study was built upon a research question - how a method to identify best practices can be statistically relevant so it's findings can be generalized and at the same time be adaptable, changing gradually its instrument to adapt to the new practices that emerges without losing previous data?

In order to answer these question, this study covered the literature on benchmarking of best practices, executed a test application of IRT on a best practices assessment questionnaire, and proposed a method to assess best practices that address most relevant critique on the topic.

The adopted methodological procedures for literature review consisted in running bibliometric analysis using two computer programs that allowed identify the most relevant articles and run in-depth analysis for the most relevant articles identified. The overall resulting relevant articles identified seemed to be well balanced in terms of novelty and relevance, with more than 17% of them from earlier than 2010. The methods applied seemed to ease the identification of relevant articles and clusters of research studies, allowing the identification of mainstream research alongside with new emerging trends.

The analysis of scientific output, presented in the first article of this compendium (Session 3), substantiates the increasing in production of articles in this research topic. 47% of the articles in the dataset corresponds to the production of the last five years (between 2010 and 2015), and the year of 2015 alone represents more than 10% of all papers in the dataset.

Despite this increasing amount of studies been published related to benchmarking of best practices, there are relevant critique that must be acknowledge. The critique put in doubt the methods adopted and go further questioning the very existence of practices that can be recognized as "best".

To counter the critiques on benchmarking studies, there are new DEA studies that propose novel methodologies that may address with statistical soundness the relation of best practices and companies' superior performance.

The IRT was proposed to address some critiques on the research of best practices because it permits flexibility to change the questionnaire with new practices without losing its capability to compare units.

The analysis, presented in the second article – session 0, was carried with a best practices assessment questionnaire with 46 items that had a database of 302 responses. The questionnaire is used by IEL/SC to assess productivity and competitiveness of small and midsized industrial companies.

The IRT model adopted was the three-parameter logistic model. The answers in the database were dichotomised because of data restriction. By the errors presented in the calibration of the items, it was possible to assume that the parameters were successfully estimated with a good margin of certainty.

The results of the IRT application emphasised the process to validate que questionnaire and describe in more depth the reasoning behind the analysis, proposing novel procedures and rationale to analyse the items of the instrument and identify aspects that can be improved in the design of questionnaires.

The *Best Practices Identification Framework* presented in the third article, in session 5, emerged from the analysis of the studies identified in the literature review.

The framework test demonstrated its applicability and opportunities, contributing to the discussion of the usage of statistical tools, especially the IRT, to the assessment of best practices and benchmarking purposes.

The most relevant critique on best practice assessment were addressed by the proposed method substantiated by methodology analysis and combining procedures that were already tested and validated by previous relevant studies.

Perhaps the main opportunity that the IRT presents is the possibility to gradually change the questionnaire without losing previous collected data. In this way, IRT permits not only to enhance items, but also to include items into the benchmarking instrument to access new practices that may emerge in the future, maintaining the test score scale and ultimately allowing the comparison of companies that have answered different questionnaires.

6.1. FUTURE RESEARCH

It is proposed that future researches address the limitations acknowledged in this study and explore the gaps that could not be covered.

The application of the IRT considered the three-parameter logistic model due to the data available that were not sufficient to run a model more appropriate for *Likert* items. Future research can apply other IRT models and further discuss the insights that it can offer to the instrument analysis.

Only some steps of the proposed framework were tested. It is necessary to run the steps 6 and 9 alongside with IEL/SC or other case study to test the framework completely. These further testing will allow to comprehend how the IRT score should respond to the expected increase of the latent trait of the average company.

The test of step 6 – "To analyse data and verify whether each good practice assessed is related to a superior performance" was done with simple correlation analysis. There is an opportunity to test DEA application in the data available.

6.2. FINAL MESSAGE

Although the relevant acknowledged critique on the research of best practices, substantiated in good grounds, one may not ignore that there are organizations that are more efficient than others – the DEA studies are a strong evidence of that.

If there are organizations that are more efficient, trying to understand what the benchmarks do to sustain their superior performance is inevitable and certainly motivates the research in the field.

The motives that drives a researcher to look for a better methodological approach are not different from the motives that drive a company to look for a better practice, both are trying to excel its own routines, and, if one can identify a state of the art method for a particular field in a particular moment, certainly it is also possible to identify a better practice for a particular desirable result in a particular moment – it is just a matter of stablishing a method that is good enough for that objective.

Moreover, the challenge that the research on best practices, and its potential results, is certainly more like a motivator than suppressor for the one who looks for answers on this topic.

Ultimately, perhaps, believing in the existence of better way for doing what we do is an ideological issue that may have driven human kind since the beginning.

7. REFERENCES

AIGNER, D. J.; CAIN, G. G. Statistical theories of discrimination in labor markets. **Industrial and Labor relations review**, p. 175–187, 1977.

ALEXANDRE, J. W. C. et al. **Uma proposta de análise de um construto para medição dos fatores críticos da gestão pela qualidade por intermédio da Teoria da Resposta ao ItemGestão & Produção**, 2002a.

ALEXANDRE, J. W. C. et al. Uma proposta de análise de um construto para medição dos fatores críticos da Gestão pela Qualidade por intermédio da Teoria da Resposta ao Item. **Gestão e Produçao**, v. 9, n. 2, p. 129–141, 2002b.

ALLEN, R. et al. Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. **Annals of Operations Research**, v. 73, p. 13–34, 1997.

AMADO, C. A. F.; SANTOS, S. P.; SEQUEIRA, J. F. C. Using Data Envelopment Analysis to support the design of process improvement interventions in electricity distribution. **European Journal of Operational Research**, v. 228, n. 1, p. 226–235, 2013.

ANDERSEN, P.; PETERSEN, N. C. A procedure for ranking efficient units in data envelopment analysis. **Management science**, v. 39, n. 10, p. 1261–1264, 1993.

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. **Teoria da Resposta ao Item: conceitos e aplicações**. São Paulo: ABE - Associação Brasileira de Estatística, 2000a.

ANDRADE, D. F. DE; TAVARES, H. R.; VALLE, R. DA C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. [s.l: s.n.].

ARROWSMITH, J.; SISSON, K.; MARGINSON, P. What can " benchmarking " offer the open method of co-ordination? **Journal of European Public Policy**, v. 11, n. 2, p. 311–328, 2004.

BALBIM JUNIOR, A.; BORNIA, A. C. Proposta de um instrumento de medida para avaliar a satisfação de clientes de bancos utilizando a Teoria da Resposta ao Item. **Gestão e Produçao**, v. 18, n. 3, p. 541–554, 2011.

BANKER, R. D.; CHARNES, A.; COOPER, W. W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. **Management Science**, v. 30, n. 9, p. 1078–1092, 1984.

BARNEY, J. Firm Resources and Sustained Competitive Advantage. **Journal of Management**, v. 17, n. 1, p. 99–120, 1991.

BERGER, A. N.; HUMPHREY, D. B. Efficiency of financial institutions: International survey and directions for future research. **European journal of operational research**, v. 98, n. 2, p. 175–212, 1997.

BOGERS, T. et al. Using citation analysis for expert retrieval in workgroups. **Proceedings of the 8th Belgian-Dutch Information Retrieval Workshop (DIR 2008)**, p. 21–28, 2008.

BOUSSOFIANE, A.; DYSON, R. G.; THANASSOULIS, E. Applied data envelopment analysis. **European Journal of Operational Research**, v. 52, n. 1, p. 1–15, 1991.

CAMP, R. C. **Benchmarking-The Search For Industry Best Practices That Lead To Superior Performance**MilwaukeeASQC Quality Press, , 1989.

CARROLL, R. J.; PRIMO, D. M.; RICHTER, B. K. Using Item Response Theory to Improve Measurement in Strategic Management: An Application to CSR. **Academy of Management Proceedings**, v. 2014, n. 1, p. 12215–12215, 20 out. 2014.

CASTRO, V. F. **Análise descritiva da reformulação da ferramenta de benchmarking para micro e pequenas empresas do Instituto Euvaldo Lodi de Santa Catarina**FlorianópolisCentro de Ciências da Administraçao e Socio-Econômicas - ESAG/UDESC, , 2010.

CAVES, D. W.; CHRISTENSEN, L. R.; DIEWERT, W. E. The economic theory of index numbers and the measurement of input, output, and productivity. **Econometrica: Journal of the Econometric Society**, v. 50, n. 6, p. 1393–1414, 1982.

CHARNES, A. et al. Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. **Journal of econometrics**, v. 30, n. 1, p. 91–107, 1985.

CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. **European Journal of Operational Research**, v. 2, n. 6, p. 429–444, 1978.

CHILINGERIAN, J. A.; DAVID SHERMAN, H. Benchmarking physician practice patterns with DEA: A multi-stage approach for cost containment. **Annals of Operations Research**, v. 67, n. 1, p. 83–116, 1996.

COELLI, T. J. et al. **An introduction to productivity and efficiency analysis**. New York: Springer Science, 1998.

COHEN, W. M.; LEVINTHAL, D. A. Absorptive Capacity: A New Perspective on Learning and Innovation. **Administrative Science Quarterly**, v. 35, n. 1, p. 128–152, 1990.

COLLINS, J. C. **Good to great: Why some companies make the leap... and others don't**. [s.l.] Radom House, 2001.

COOK, W. D.; TONE, K.; ZHU, J. Data envelopment analysis: Prior to choosing a model. **Omega**, v. 44, p. 1–4, 2014.

COOPER, R. G. The Dimensions of Industrial New Product Success and Failure. **Journal of Marketing**, v. 43, n. 3, p. 93–103, 1979.

COOPER, R. G.; EDGETT, S. J.; KLEINSCHMIDT, E. J. New product portfolio management: practices and performance. **Journal of product innovation management**, v. 16, n. 4, p. 333–351, 1999.

COOPER, R. G.; EDGETT, S. J.; KLEINSCHMIDT, E. J. Benchmarking best NPD practices-I. **Research-Technology Management**, v. 47, n. 1, p. 31–43, 2004a.

COOPER, R. G.; EDGETT, S. J.; KLEINSCHMIDT, E. J. Benchmarking Best NPD Practices - II. **Research-Technology Management**, v. 47, n. 1, p. 50–59, 2004b.

COSTA, D. B. et al. Benchmarking Initiatives in the Construction Industry : Lessons Learned and Improvement Opportunities. **Journal of Management in Engineering**, v. 22, n. 4, p. 158–167, 2007.

COUGHLAN, P.; COGHLAN, D. Action research for operations

management. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 220–240, 2002.

CRESWELL, J. W. Research Design. **SAGE Publications**, 2009.

DAI, X.; KUOSMANEN, T. Best-practice benchmarking using clustering methods: Application to energy regulation. **Omega**, v. 42, n. 1, p. 179–188, 2014.

DATTAKUMAR, R.; JAGADEESH, R. A review of literature on benchmarking. **Benchmarking: An International Journal**, v. 10, n. 3, p. 176–209, 2003.

DAVIES, A. J.; KOCHHAR, A. K. Manufacturing best practice and performance studies: a critique. **International Journal of Operations & Production Management**, v. 22, n. 3, p. 289–305, 2002.

DELBRIDGE, R.; LOWE, J.; OLIVER, N. The process of benchmarking: a study from the automotive industry. **International Journal of Operations & Production Management**, v. 15, n. 4, p. 50–62, 1995.

DEMING, W. E. **Quality, productivity, and competitive position**. [s.l.] Massachusetts Institute of Technology Center for Advanced En, 1982.

DESPOTIS, D. K. Measuring human development via data envelopment analysis: the case of Asia and the Pacific. **Omega**, v. 33, p. 385–390, 2005.

DESPOTIS, D. K.; SMIRLIS, Y. G. Data envelopment analysis with imprecise data. **European Journal of Operational Research**, v. 140, n. 1, p. 24–36, 2002.

DOYLE, J.; GREEN, R. Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. **Journal of the operational research society**, v. 45, n. 5, p. 567–578, 1994.

DREW, S. A. W. From knowledge to action: the impact of benchmarking on organizational performance. **Long Range Planning**, v. 30, n. 3, p. 427–441, 1997.

DYSON, R. G. et al. Pitfalls and protocols in DEA. **European Journal of operational research**, v. 132, n. 2, p. 245–259, 2001.

DYSON, R. G.; THANASSOULIS, E. Reducing weight flexibility in data envelopment analysis. **Journal of the Operational Research Society**, v. 39, p. 563–576, 1988.

EGAN, J. **Rethinking Construction, Construction Task Force Report for Department of the Environment, Transport and the Regions**LondonHMSO, , 1998.

EISENHARDT, K. M. Building Theories from Case Study Research. **Academy of Management Review**, v. 14, n. 4, p. 532–550, 1989.

EL-MASHALEH, M. S.; EDWARD MINCHIN JR, R.; O'BRIEN, W. J. Management of construction firm performance using benchmarking. **Journal of Management in Engineering**, v. 23, n. 1, p. 10–17, 2007.

EVANS, M. T. P.; TISAK, D. J.; WILLIAMSON, D. F. Twenty-first century benchmarking: searching for the next generation. **Benchmarking: An International Journal**, v. 19, n. 6, p. 760–780, 2012.

FARRELL, M. J. The Measurement of Productive Efficiency. **Journal of the Royal Statistical Society. Series A (General)**, v. 120, n. 3, p. 253–290, 1957.

FOLZ, D. H. Service Quality and Benchmarking the Performance of Municipal Services. **Public Administration Review**, v. 64, n. 2, p. 209–220, 2001.

FORKER, L. B.; MENDEZ, D. An analytical method for benchmarking best peer suppliers. **International Journal of Operations & Production Management**, v. 21, n. 1, p. 195–209, 2001.

FRANCIS, G.; HOLLOWAY, J. What have we learned? Themes from the literature on best-practice benchmarking. **International Journal of Management Reviews**, v. 9, n. 3, p. 171–189, 2007.

GAREY, M. R.; JOHNSON, D. S. **Computers and intractability: a guide to the theory of NP-completeness. San Francisco, LA: Freeman**, 1979.

GIL, A. C. Métodos e técnicas de pesquisa social. In: **Métodos e técnicas de pesquisa social**. [s.l.] Atlas, 2010.

GOLDBERG, D. E.; HOLLAND, J. H. Genetic algorithms and machine learning. **Machine learning**, v. 3, n. 2, p. 95–99, 1988.

GRIFFIN, A. PDMA Research on New Product Development Practices: Updating Trends and Benchmarking Best Practices. **Journal of Product Innovation Management**, v. 14, n. 6, p. 429–458, 1997.

GUPTA, S.; VERMA, R.; VICTORINO, L. Empirical research published in production and operations management (1992–2005): trends and future research directions. **Production and Operations Management**, v. 15, n. 3, p. 432–448, 2006.

HALL, D. J. The role of creativity within best practice manufacturing. **Technovation**, v. 16, n. 3, p. 115–121, 1996.

HAMMER, M. **Reengineering the Corporation**. [s.l.] New york: Harper collins, 1993.

HANSON, P.; VOSS, C. Benchmarking best practice in European manufacturing sites. **Business Process Management Journal**, v. 1, n. 1, p. 60–74, 1995.

HARRINGTON, J. H. **The fallacy of universal best practicesTotal Quality Management & Business Excellence**, 2004. Disponível em: <http://proxy1.ncu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=13837820&site=ehost-live>

HAYES, R. H.; WHEELWRIGHT, S. C. **Restoring our competitive edge: competing through manufacturing**New YorkJohn Wiley & Sons, , 1984.

HORTA, I. M.; CAMANHO, A. S.; MOREIRA DA COSTA, J. Performance assessment of construction companies: A study of factors promoting financial soundness and innovation in the industry. **International Journal of Production Economics**, v. 137, n. 1, p. 84–93, 2012.

IEL/SC. **Questionário do Benchstar**Instituto Euvaldo Lodi de Santa Catarina - IEL/SC, , 2010.

JURAN, J. M. **Quality and Income**. [s.l.] McGraw-Hill, 1999.

KAHN, K. B.; BARCZAK, G.; MOSS, R. Perspective: establishing an NPD best practices framework. **Journal of Product Innovation Management**, v. 23, n. 2, p. 106–116, 2006.

KIRKPATRICK, S. et al. Optimization by simulated annealing. **Science**, v. 220, n. 4598, p. 671–680, 1983.

KNUTAS, A. et al. **Cloud-Based Bibliometric Analysis Service for Systematic Mapping Studies**. CompSysTech '15 Proceedings of the 16th International Conference on Computer Systems and Technologies. **Anais**...New York: ACM, 2015

KORPELA, J.; TUOMINEN, M. Benchmarking Logistics Performance with an Application of the Analvtic Hierarchv Process. **IEEE Transactions on Engineering Management**, v. 43, n. 3, p. 323–333, 1996.

KUMAR, S.; CHANDRA, C. Enhancing the effectiveness of benchmarking in manufacturing organizations. **Industrial Management & Data Systems**, v. 101, n. 2, p. 80–89, 2001.

KUULA, M.; PUTKIRANTA, A. Longitudinal benchmarking studies in operations management: lessons learned. **Benchmarking: An International Journal**, v. 19, n. 3, p. 358–373, 2012.

KUULA, M.; PUTKIRANTA, A.; TOIVANEN, J. Coping with the change: a longitudinal study into the changing manufacturing practices. **International Journal of Operations & Production Management**, v. 32, n. 2, p. 106–120, 2012.

LAUGEN, B. T. et al. Best manufacturing practices: What do the best-performing companies do? **International Journal of Operations & Production Management**, v. 25, n. 2, p. 131–150, 2005.

LEE, S.-H.; THOMAS, S. R.; TUCKER, R. L. Web-based benchmarking system for the construction industry. **Journal of Construction Engineering and Management**, v. 131, n. 7, p. 790–798, 2005.

LEE, S. H. et al. Economic value of combined best practice use. **Journal of Management in Engineering**, v. 21, n. 3, p. 118–124, 2005.

LEMA, N. M.; PRICE, A. D. F. Benchmarking: Performance improvement

toward competitive advantage. **Journal of Management in Engineering**, v. 11, n. 1, p. 28–37, 1995.

LESEURE, M. J. et al. Adoption of promising practices : A systematic review of the evidence. **International Journal of Management Reviews**, v. 5, n. 3, p. 169–190, 2004.

MACKENZIE, S.; PODSAKOFF, P.; PODSAKOFF, N. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. **MIS quarterly**, v. 35, n. 2, p. 293–334, 2011.

MEEUSEN, W.; DEN BROECK, J. Efficiency estimation from Cobb-Douglas production functions with composed error. **International economic review**, v. 18, n. 2, p. 435–444, 1977.

MENKE, M. M. Making R&D Portfolio Management More Effective. **Research Technology Management**, v. 56, n. 5, p. 34–44, 2013.

MORAIS, P.; CAMANHO, A. S. Evaluation of performance of European cities with the aim to promote quality of life improvements. **Omega**, v. 39, n. 4, p. 398–409, 2011.

MOREIRA JUNIOR, F. DE J. Aplicações da teoria da resposta ao item (TRI) no Brasil. **Rev. Bras. Biom**, v. 28, n. 4, p. 137–170, 2010.

OZER, M.; CHEN, Z. Do the best new product development practices of US companies matter in Hong Kong? **Industrial Marketing Management**, v. 35, n. 3, p. 279–292, 2006.

PACHECO, J. A.; ANDRADE, D. F. DE; BORNIA, A. C. Benchmarking by Item Response Theory (BIRTH): a benchmarking method using IRT to build competitiveness scales for Brazilian technology higher education. **Benchmarking: An International Journal**, v. 22, n. 5, p. 945–962, 2015.

PAGE, L. et al. **The PageRank Citation Ranking: Bringing Order to the Web**Stanford InfoLab, , 1998. Disponível em: <http://ilpubs.stanford.edu:8090/422>

PHILLIPS, P.; APPIAH-ADU, K. Benchmarking to Improve the Strategic Planning Process in the Hotel Sector. **Service Industries Journal**, v. 18, n.

l, p. 1–17, 1998.

PORTER, M. E. **Competitive strategy: Techniques for analyzing industries and competition**New YorkThe Free Press, , 1980.

RAGATZ, G. L.; HANDFIELD, R. B.; SCANNELL, T. V. Success Factors for Integrating Suppliers into New Product Development. **Journal of Product Innovation Management**, v. 14, n. 3, p. 190–202, 2003.

RASCH, G. **Probabilistic models for some intelligence and attainment tests**. Chicago: MESA Press, 1993.

RUIZ, J. L.; SEGURA, J. V.; SIRVENT, I. Benchmarking and target setting with expert preferences: An application to the evaluation of educational performance of Spanish universities. **European Journal of Operational Research**, v. 242, n. 2, p. 594–605, 2015.

SAATY, T. L. **The analytic hierarchy process: planning, priority setting, resources allocation**. New York: McGraw, 1980.

SCHIELE, H.; KRUMMAKER, S. Consortium benchmarking: Collaborative academic–practitioner case study research. **Journal of Business Research**, v. 64, n. 10, p. 1137–1145, 2011.

SCHONBERGER, R. J. **World class manufacturing: the lessons of simplicity applied**New YorkThe Free Press, , 1986.

SHERMAN, H. D. Hospital Efficiency Measurement and Evaluation: Empirical Test of a New Technique. **Medical care**, v. 22, n. 10, p. 922–938, 1984.

SHERMAN, H. D.; LADINO, G. Managing Bank Productivity Using Data Envelopment Analysis (DEA). **Interfaces**, v. 25, n. 2, p. 60–73, 1995.

SISKOS, Y.; DESPOTIS, D. K.; GHEDIRI, M. Multiobjective modelling for regional agricultural planning: case study in Tunisia. **European Journal of Operational Research**, v. 77, n. 3, p. 375–391, 1994.

SMIRLIS, Y. G.; MARAGOS, E. K.; DESPOTIS, D. K. Data envelopment analysis with missing values: An interval DEA approach. **Applied Mathematics and Computation**, v. 177, n. 1, p. 1–10, 2006.

SOUSA, R.; VOSS, C. Contingency research in operations management practices. **Journal of Operations Management**, v. 26, n. 6, p. 697–713, nov. 2008.

SPENDOLINI, M. J. **The benchmarking book**. [s.l.] Amacom New York, NY, 1992.

STOCKING, M. L.; LORD, F. M. Developing a Common Metric in Item Response Theory. **Applied Psychological Measurement**, v. 7, n. 2, p. 201–210, 1 abr. 1983.

STRAUB, D. W. Validating instruments in MIS research. **MIS quarterly**, v. 13, n. 2, p. 147–169, 1989.

SUAREZ, E.; CALVO-MORA, A.; ROLDÁN, J. L. The role of strategic planning in excellence management systems. **European Journal of Operational Research**, v. 248, n. 2, p. 532–542, 2016.

SZULANSKI, G. Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm. **Strategic Management Journal**, v. 17, n. 2, p. 27–43, 1996.

TONE, K. A slacks-based measure of efficiency in data envelopment analysis. **European journal of operational research**, v. 130, n. 3, p. 498–509, 2001.

TONGZON, J. Efficiency measurement of selected Australian and other international ports using data envelopment analysis. **Transportation Research Part A: Policy and Practice**, v. 35, n. 2, p. 107–122, 2001.

TRIERWEILLER, A. C. et al. Environmental management performance for Brazilian industrials: measuring with the item response theory. **Work: A Journal of Prevention, Assessment and Rehabilitation**, v. 41 Suppl 1, p. 2179–86, jan. 2012.

ULUSOY, G.; IKIZ, I. Benchmarking best manufacturing practices: a study into four sectors of Turkish industry. **International Journal of Operations & Production Management**, v. 21, n. 7, p. 1020–1043, 2001.

VAN ECK, N. J.; WALTMAN, L. Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, v. 84, n. 2, p.

523–538, 2010.

VARGAS, V. C. C. et al. Avaliação dos intangíveis: uma aplicação em capital humano. **Gestão e Produçao**, v. 15, n. 3, p. 619–634, 2008.

VASCONCELOS, A. P.; LEZANA, Á.; ANDRADE, D. F. Escala para medição do sucesso de micro e pequenas empresas de Santa Catarina com o apoio da teoria da resposta ao item. **Revista Brasileira de Gestão e Desenvolvimento Regional**, v. 9, n. 1, p. 91–119, 2013.

VOSS, C. et al. **The Competitiveness of European Manufacturing - A Four Country StudyBusiness Strategy Review**, 1995. Disponível em: <http://doi.wiley.com/10.1111/j.1467-8616.1995.tb00169.x>

VOSS, C. A. Operations Management- from Taylor to Toyota- and Beyond? **British Journal of Management**, v. 6, n. Special Issue, p. S17–S29, 1995.

VOSS, C. A.; AHLSTROM, P.; BLACKMON, K. Benchmarking and operational performance: some empirical results. **Internation Journal of Operations & Production Management**, v. 17, n. 10, p. 1046–1058, 1997.

VOSS, C.; ÅHLSTRÖM, P.; BLACKMON, K. Benchmarking and operational performance: some empirical results. **Benchmarking: An International Journal**, v. 4, n. 4, p. 273–285, 1997.

VOSS, C.; CHIESA, V.; COUGHLAN, P. Developing and Testing Benchmarking and Self-assessment Frameworks in Manufacturing. **International Journal of Operations & Production Management**, v. 14, n. 3, p. 83–100, 1994.

VOSS, C.; TSIKRIKTSIS, N.; FROHLICH, M. Case research in operations management. **International Journal of Operations & Production Management**, v. 22, n. 2, p. 195–219, 2002.

WANG, C. Conscious Capitalism Firms: Do They Behave As Their Proponents Say? **California Management Review**, v. 55, n. 3, p. 60–87, 2013.

WATSON, G. H. **Strategic benchmarking: How to rate your company's performance against the world's best**. New York: Wiley, 1993.

WELLSTEIN, B.; KIESER, A. Trading “best practices”--a good practice? **Industrial and Corporate Change**, v. 20, n. 3, p. 683–719, 2011.

WILLIAMS, J.; BROWN, C.; SPRINGER, A. Overcoming benchmarking reluctance: a literature review. **Benchmarking: An International Journal**, v. 19, n. 2, p. 255–276, 2012.

WOMACK, J. P.; JONES, D. T.; ROOS, D. **The machine that changed the world**. New York: Simon and Schuster, 1990.

YARROW, D.; HANSON, P.; ROBSON, A. **Made in the 21st Century: How Far Have we Come on the Journey to Excellence?Total Quality Management & Business Excellence**, 2004.

YASIN, M. M. The theory and practice of benchmarking: then and now. **Benchmarking: An International Journal**, v. 9, n. 3, p. 217–243, 2002.

ZIMOWSKI, M. et al. **Bilog-MG**LincolnwoodScientific Software International, Inc., , 2003.